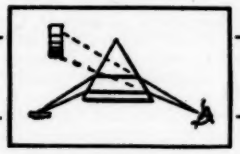
Journal of the OPTICAL SOCIETY *of* AMERICA

VOLUME 38
No. 11



NOVEMBER
1948

Control of Photographic Printing: Improvement in Terminology and Further Analysis of Results . . . . . LOYD A. JONES AND C. N. NELSON 897

On the Theory of Imperfect Diffraction Gratings . . . . . U. FANO 921

Transmission Filters for the Ultraviolet . . . . . M. KASHA 929

Differential Adjustment Method of Refining Optical Systems . . . WALDEMAR M. STEMPEL 935

Reflection and Transmission of Light by Any System of Parallel Isotropic Films . A. W. CROOK 954

Coefficients of Reflection in the Ultraviolet of Magnesium Carbonate and Oxide . . . . . FRANK BENFORD, SALLY SCHWARZ, AND OWEN P. LLOYD 964

July 1948 Meeting of the International Commission of Optics 965

Light Scattering in Optical Glass . . . . . D. H. RANK AND A. E. DOUGLAS 966

Illuminant-Stable Color Vision Test. II. Experimental-Statistical Evaluation . . . . . ELLIS FREEMAN AND M. A. ZACCARIA 971

Short Wave-Length Cut-Off Filters for the Ultraviolet . . . . . ARNOLD M. BASS 977

Notes on the $\cos^4$ Law of Illumination . . . . . MAX REISS 980

Letters to the Editor:

Elimination of Water Vapor in Infra-Red Spectrometers . . . . . PAUL A. GIGUÈRE AND RICHARD M. BADGER 987

Infra-Red Absorption of Deposited Blacks . . . . . EARLE K. PLYLER AND JOSEPH J. BALL 988

Emissivity of Globar . . . . . SHIRLEIGH SILVERMAN 989

Book Review 989

Published monthly
at Prince and Lemon Sts., Lancaster, Pa.
by the
AMERICAN INSTITUTE OF PHYSICS
Incorporated
57 East 55 Street, New York 22, New York
FOR AND UNDER THE EDITORIAL SUPERVISION OF THE
OPTICAL SOCIETY OF AMERICA

THE OPTICAL SOCIETY OF AMERICA

Officers of the Society

RUDOLF KINGSLAKE, *President*
Eastman Kodak Company
Rochester, New York

WILLIAM F. MEGGERS, *Vice President*
National Bureau of Standards
Washington 25, D. C.

ARTHUR C. HARDY, *Secretary*
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

W. W. GRAEPER, *Treasurer*
Bausch & Lomb Optical Company
Rochester 2, New York

The Board of Directors consists of the above officers; the Editor of the Journal; the Past President; and five directors-at-large.

G. R. HARRISON, *Editor and Past President*
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

Directors-at-Large

FRANK BENFORD
General Electric Company
Schenectady, New York

F. M. BISHOP
Eastman Kodak Company
Rochester 4, New York

WALLACE R. BRODE
Ohio State University
Columbus 10, Ohio

A. H. BENNETT
American Optical Company
Buffalo 11, New York

E. H. LAND
Polaroid Corporation
Cambridge, Massachusetts

Local Sections

STANLEY S. BALLARD, *Secretary for the Local Sections*

Rochester Section

JOHN H. MCLEOD, *President*
HARRY D. POLSTER, *Vice President*
JAMES R. BENFORD, *Secretary*
LUVILLE T. STEADMAN, *Treasurer*

Detroit Section

DAVID L. FRY, *Chairman*
KARL W. BEYER, *Vice Chairman*
DANIEL L. HARMON, *Secretary*
W. R. O'NEILL, *Treasurer*

Niagara Frontier Section

GALEN PORTER, *President*
LELAND E. HAVENS, *Secretary*
CLAIR M. BIRDSALL, *Treasurer*

Ohio Valley Section

M. E. MERCHANT, *Chairman*
J. F. WOODRUFF, *Chairman-elect*
C. S. MILLS, *Secretary-Treasurer*

Patrons

AMERICAN OPTICAL COMPANY — BAUSCH & LOMB OPTICAL COMPANY — EASTMAN KODAK COMPANY

Corporation Members

ALLIED CHEMICAL AND DYE CORPORATION, NATIONAL ANILINE DIVISION, Buffalo, New York
AMERICAN CYSTOSCOPE MAKERS, INC., New York, New York
AMERICAN OPTICAL COMPANY, Southbridge, Massachusetts
ARGUS, INCORPORATED, Ann Arbor, Michigan
BAUSCH & LOMB OPTICAL COMPANY, Rochester, New York
BELL & HOWELL COMPANY, Chicago, Illinois
BELL TELEPHONE LABORATORIES, INC., New York, New York
CENTURY LIGHTING, INC., New York, New York
CORNING GLASS WORKS, Corning, New York
R. R. DONNELLEY & SONS COMPANY, Chicago, Illinois
E. I. DUPONT DE NEMOURS & COMPANY, INC., Chemical Department, Experimental Station, Wilmington, Delaware
EASTMAN KODAK COMPANY, Rochester, New York
EDMUND SALVAGE COMPANY, Oaklyn-Audubon, New Jersey
FAIRCHILD AVIATION CORPORATION, Jamaica, New York
FARRAND OPTICAL COMPANY, INC., New York, New York
FISH-SCHURMAN CORPORATION, New York, New York
GAERTNER SCIENTIFIC CORPORATION, Chicago, Illinois
W. & L. E. GURLEY, Troy, New York
INTERCHEMICAL CORPORATION, New York, New York
KEUFFEL & ESSER COMPANY, Hoboken, New Jersey
KOLLSMAN INSTRUMENT DIVISION OF SQUARE D COMPANY, Elmhurst, New York
NATIONAL CARBON COMPANY, Cleveland, Ohio
NATIONAL TECHNICAL LABORATORIES, South Pasadena, California
PERKIN-ELMER CORPORATION, Glenbrook, Connecticut
POLAROID CORPORATION, Cambridge, Massachusetts
PORTMAN MACHINE TOOL COMPANY, New Rochelle, New York
SIMPSON OPTICAL MANUFACTURING COMPANY, Chicago, Illinois
SOCONY-VACUUM LABORATORIES, Paulsboro, New Jersey
ARTHUR H. THOMAS COMPANY, Philadelphia, Pennsylvania
UNIVIS LENS COMPANY, Dayton, Ohio
ZENITH OPTICAL COMPANY, Huntington, West Virginia

THE OPTICAL SOCIETY OF AMERICA was organized in 1916 "*to increase and diffuse the knowledge of optics in all its branches, pure and applied, to promote the mutual interests of investigators of optical problems, of designers, manufacturers and users of optical instruments and apparatus of all kinds and to encourage cooperation among them.*" It aims to serve the interests of all who are engaged in any branch of optics from fundamental research to the manufacture of optical goods.

MEMBERSHIP

Regular Members: "Any person who has contributed materially to the advancement of optics is eligible for regular membership."

Associate Members: "Any person who is interested in optics is eligible for associate membership."

Corporate Members: "Any corporation desirous of encouraging the work of the Society is eligible for corporation membership."

Patrons: The payment, within any one year, of $1000 or more toward the expenses of the Society by any corporation or individual entitles such corporation or individual to the rank of patron.

The Society cordially invites to membership all who are interested in any branch of optics, either in research, in instruction, in optical or illuminating engineering, in the manufacture and distribution of optical goods of all kinds, or in physiological and medical optics.

ANNUAL DUES

Regular members, $8.50; Associate members, $6.00; Corporate members, $50.00. These dues include subscription to the Society's official journal, the *Journal of the Optical Society of America.*

CHANGES OF ADDRESS

All changes of address of members of the Optical Society of America should be communicated to the Secretary, Professor Arthur C. Hardy, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 38, NUMBER 11 NOVEMBER, 1948

Control of Photographic Printing: Improvement in Terminology and Further Analysis of Results*

LOYD A. JONES AND C. N. NELSON
Eastman Kodak Company, Rochester 4, New York
(Received August 18, 1948)

Further study is made of the statistical correlation between the *density scales* of negatives and the *sensitometric exposure scales* of photographic papers. A formula relating these two variables is proposed, for use as a guide in photographic printing, whereby the density scale of the negative can be used as a means of choosing the grade of paper to be used in making the print. The suggestions of Sanders regarding statistical methods of analyzing the data are discussed. The term *useful exposure scale* is abandoned and the terms *sensitometric exposure scale* and *transition point index* are proposed. It is suggested that the sensitometric exposure scale be adopted as the basis for grading photographic papers and that the new numbers be called *scale indices*. Reasons are given why the *contrast* of photographic paper is not suitable as a basis for deriving grade numbers. For specifying the shapes of the $D-\log E$ curves of photographic papers, the method proposed by Morrison is recommended. Data are presented on the frequency of occurrence of the various density scales of amateur negatives.

I. INTRODUCTION

IN a previous communication[1] we presented a large volume of experimental data pertaining to the measurable characteristics of photographic negatives and printing papers which, when used together, gave prints of excellent quality. The purpose of this work was to find (*i*) a sensitometric method for determining log exposure scales and speeds of photographic papers which are significant in practical photographic printing, and (*ii*) a method of controlling photographic printing in which the log exposure scales and speeds of the papers are related to the densities in the negatives so that the optimum grade of paper for printing each negative and the optimum printing exposure can be predetermined, thus eliminating the necessity of making trial prints.

Prints were made from 170 negatives of different scenes using five or more different printing exposures on each of three and sometimes four grades of paper for each negative. Consequently, the prints varied systematically in contrast and density. These prints were examined by judges working separately who selected the most pleasing print (referred to in the following pages as *first-choice* prints) for each negative and thus determined the optimum grade of paper for each negative and the optimum printing exposure. The maximum and minimum densities in the negatives and in the chosen prints were measured and located on the $D-\log E$ curves of the negative and positive materials.

* Communication No. 1199 from the Kodak Research Laboratories.

[1] L. A. Jones and C. N. Nelson, "The control of photographic printing by measured characteristics of the negative," J. Opt. Soc. Am. **32**, 558 (1942).

As a result of this work a *sensitometric criterion* was proposed for determining the log exposure scales of photographic papers. Because of its significance in practical photographic printing and its good correlation with prints of high photographic quality, this criterion was adopted by the American Standards Association[2] in February, 1947. The criterion defines a log exposure interval lying between two limiting points on the density-log exposure curve of the paper, one point being at a gradient of $0.1\bar{G}$ on the toe and the other at a gradient of $1.0\bar{G}$ on the shoulder of the curve, where $\bar{G}$ is the average gradient measured between these two points. It was shown in our previous communication that this criterion gave log exposure scale values which, on the average, corresponded closely to the portions of the $D-\log E$ curves used in making the *first-choice* prints.

The correlation between the density scales, DS_n, of the negatives and the sensitometric log exposure scales, $\log SES_y$, of the papers in the case of prints of *first-choice* quality was not perfect. It was found, however, that by a relatively small sacrifice in the average quality of the entire group of 170 prints a unique relationship between DS_n and $\log SES_y$ could be established. Thus, a method for choosing the proper grade of paper for the printing of a negative of known density scale was found which proved to give results sufficiently good for practical application.

Since the publication of that communication,[1] further data have been accumulated on the relation between the density scales of negatives and the log exposure scales of papers which yield prints of excellent quality. It has become increasingly evident that there are several aspects of exposure scale which have not been properly distinguished from each other by suitable terminology. We wish, therefore, at this time to discuss these aspects of exposure scale and suggest a nomenclature which distinguishes clearly and unequivocally between them. In particular, we wish to propose the use of the term *sensitometric exposure scale*, SES_y, for the values derived by use of the sensitometric criterion previously described. The term *useful exposure scale*, which was originally applied to the average of the density scales of the negatives giving *first-choice* prints on each grade of paper, has been found to be confusing when applied to a photographic paper, since its numerical value depends upon the frequency distribution of the density scales of the negatives used and upon the spacing of the grades of paper used in making the prints, and hence cannot be a *unique* characteristic of any one paper by itself. Therefore, we propose that the term *useful exposure scale* be abandoned.

A paper by Sanders[3] has been published recently in which he seeks to explain an interesting empirical relation between the negatives and positives which we derived from the data contained in our previous paper.[1] Our analysis of the statistical print judgment data revealed that for maximum yield of high quality prints a surprising rule should be followed: For the soft papers, the density scales of the negatives should in most cases exceed the sensitometric exposure scale of the paper, whereas, for the hard papers, the density scales of the negatives should in most cases be less than the sensitometric exposure scale of the paper. This result occurred despite the fact that the sensitometrically determined log exposure scale values were very nearly equal to the average of the density scales of the negatives which gave the *first-choice* prints on each paper. Sanders uses the methods of statistical mathematics to show that this general type of result appears not only in our final print judgment data, where a unique relation between negative density scale and paper log exposure scale was established, but also in our original print judgment data involving only *first-choice* prints. While our experimental data do not lend themselves readily to deriving quantitative conclusions using his method of approach, his treatment has contributed to a better understanding of the factors involved, and in the following pages from time to time we shall refer to his work.

Sanders suggests that a modified log exposure scale be derived from the sensitometrically determined log exposure scale as a means of simplifying, for some practical applications, the calculations involved in determining which nega-

[2] *American Standard for Sensitometry of Photographic Papers* (American Standards Association, New York, 1947), Z38, 2.3.

[3] T. D. Sanders, "The relation between the density scale of negatives and the exposure scale of papers," Phot. J. **87B**, 13 (1947).

tives should be printed on each grade of paper. We propose to call this new aspect of log exposure scale the *transition point index*, because this term suggests its intended use and will not be confused with the sensitometric exposure scale.

Section II of this paper deals with the proposals regarding *sensitometric exposure scale*, *useful exposure scale*, and *transition point index*. Other aspects of exposure scale are also discussed. The results of a redetermination of transition points by a rejudgment of our prints are given and the proposals of Sanders are discussed more fully.

In Section III we propose that the sensitometric exposure scale be used as a basis for the grading of photographic printing papers. The term *scale index* is suggested for the new values, since the term "sensitometric log exposure scale" is too long for general use. In Section IV Morrison's method for specifying the shapes of the $D-\log E$ curves of photographic papers is recommended.

Section V presents the available data regarding the frequency distribution of the density scales of negatives in amateur photography.

Section VI gives a summary and conclusions in itemized form.

II. SENSITOMETRIC EXPOSURE SCALE AND ITS RELATION TO NEGATIVE DENSITY SCALE

It will be necessary first to review a part of our previous paper and some of the conclusions drawn therefrom.

The characteristics of the group of negatives used, all made under carefully controlled conditions, and of the exterior scenes which the negatives represent, were given in detail in a communication by Jones and Condit.[4] For the present problem, our chief interest in these 170 negatives lies in the values of their density scales, DS_n, which is the difference between the maximum and minimum densities found in each negative. All these negatives were obtained with normal** camera exposures and normal development treatment. Hence, the density scale in each case it determined largely by the luminance scale, BS_o, of the object. The camera-lens equipment used in making all these negatives had optical and mechanical characteristics resulting in a relatively small amount of flare light in the camera. The flare light was therefore low and constant, with the exception of the flare light variation attributable to the variation in the distribution of luminance within the scenes photographed. In any case, the variation in the density scale among the members of this group of negatives is attributable almost wholly to the variation in the luminance scale of the scene and only in small part to differences in the extent to which development was carried or to underexposure or overexposure in making the negatives. For the making of prints from these negatives, we chose a group of six grades of Velox paper varying with respect to the logarithm of the sensitometric exposure scale, $\log SES_y$, from 0.54 to 1.78.

The problem which confronted us in obtaining the best possible prints from each of these negatives is very similar to that which confronts the amateur photographer or the commercial photo finisher who has a large number of negatives varying with respect to density scale from which he must make the best possible prints by using a *few* grades of paper differing with respect to exposure scale, ES_y.

It is well known that, in general, a negative with a low value of density scale must be printed on a positive material having a relatively short exposure scale, and vice versa. It was therefore possible to estimate within a grade or two which grade of printing paper would yield the best print from any given negative. Each negative was printed on three or four adjacent grades of the positive materials which we were using and in each case a series of prints, usually five or six, varying with respect to printing exposure, was made with the nominally correct exposure at about the middle of the exposure series. In this way 170 groups of prints, one for each negative,

[4] L. A. Jones and H. R. Condit, "The brightness scale of exterior scenes and the computation of correct photographic exposure," J. Opt. Soc. Am. **31**, 651 (1941).

** As used here, the term *normal* camera exposure serves to exclude underexposure and overexposure. If the scene element of lowest luminance is rendered by a density in the negative which is less than the density corresponding to the $0.3\bar{G}$ speed criterion for negative materials, the negative is said to be underexposed. Such a negative has not received sufficient exposure to permit the making of an excellent print. If the scene element of maximum luminance is rendered by a density on the shoulder of the characteristic curve where the gradient is less than one-half of that of the maximum gradient, γ, of the negative material, the negative is considered to be overexposed.

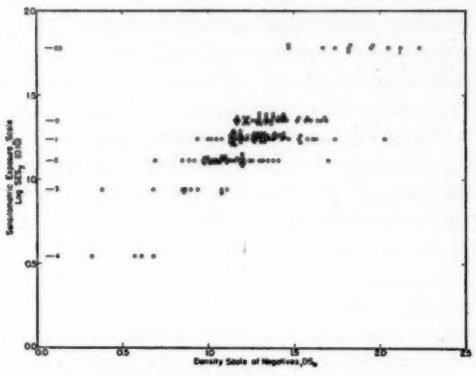


FIG. 1. Print judgment data pertaining to the first-choice prints. Relation between the sensitometric exposure scales of the papers and density scales of the negatives.

were obtained, the total number of prints being approximately 3000. All prints were made by contact. Negative densities were measured in such a way as to yield values of effective contact printing density for the positive material—light source combination used in making the prints. All prints were developed to the same extent.

Each group of prints, consisting of from 15 to 20 members, was examined under standard conditions of illumination by a group of 30 observers. Each judge worked individually and was not influenced by the opinions of other judges, since care was taken that no judge should have any knowledge whatever of the prints chosen by those who had previously examined the prints. The attendant in charge of exhibiting the prints, recording, etc., was not permitted to discuss the prints or the judgments of other observers with the individuals making the print judgments. All of these precautions were taken to insure that the judgments of quality were entirely unprejudiced.

For each group of prints from the same negative, each judge was asked to choose: first, the printing exposure yielding the best print in each series of prints on a given grade of positive material; second, the grade of printing paper which gave the print of first-choice (that is, best possible), second-choice, and third-choice qualities. The answers to the first of these questions were used in our previous paper to establish a rational method for determining the speed of photographic printing papers and a criterion for determining the printing exposure from the measured densities of the negatives. With this particular aspect of the problem we are not at present concerned.

From the answers to the second of these two questions, information was obtained on the relation between the density scale of a negative and the sensitometric exposure scales of the positive materials yielding prints of first-, second-, and third-choice qualities. Figure 1 (Fig. 28 of reference 1) shows graphically the results of the judgment data on the first-choice prints. This figure has been redrawn to give better photomechanical reproduction than was obtained in the previous communication. Each small circle represents a print of first-choice quality and the coordinates give the values of density scale of the negative, DS_n, and the logarithm of the sensitometric exposure scale, $\log SES_y$, of the positive material.

In order that there may be no misunderstanding, we shall use the term *sensitometric exposure scale*, abbreviated SES_y, to indicate that aspect of exposure scale measured by the purely sensitometric technique which was recommended in our previous communication. Thus, the sensitometric exposure scale is defined as the log exposure interval between two points, one located on the toe and the other on the shoulder of the characteristic curve. The gradient at the point on the toe is equal to $0.10\bar{G}$ and that on the shoulder is $1.0\bar{G}$, where the line connecting these two points has a slope, $\bar{G}$. This constitutes the fractional-gradient criterion which we recommended for determining the exposure scale which we now shall refer to as *sensitometric exposure scale*, not *useful exposure scale*, as was suggested in our previous paper.

The small circles representing the individual prints, in Fig. 1, should all be plotted on a single horizontal line passing through one of the $\log SES_y$ ordinate values. They are distributed above and below such lines only to show each point individually.

From an examination of Fig. 1, it is obvious that there is no *unique* relationship between DS_n and $\log SES_y$, and hence the value of negative density scale cannot be an unambiguous guide for the selection of the grade of paper to be used in making a print of first-choice quality. It is interesting, however, to examine the relationship

between the average value of the density scales of those negatives used in the making of each of these six groups of first-choice prints and the values of sensitometric exposure scale of the positive materials. This relationship is shown in Fig. 2. The correlation between the two variables is extremely good and it appears that the relationship can be represented satisfactorily by a straight line. The solid straight line has been determined by the method of least squares as giving the best possible fit with the experimental values. In the second column of Table I are given the values of $\log SES_y$.

The distance, expressed in terms of log exposure scale, between the two points on the characteristic curve of the positive material where the maximum and minimum densities of a given negative are printed must be numerically equal to the density scale of the negative. Hence, by averaging the density scale values of the negatives for one of the groups shown in Fig. 1, a value is obtained which is equal to the average value of that part of the log exposure scale of the positive material which was used in making the prints in that group. The values obtained in this manner are in the third column of Table I. We might refer to these values either as the average DS_n of the negatives used in making the prints in each of these groups, or as average log exposure scale values applying to each of the first-choice groups.

In our previous communication we suggested the term *useful log exposure scale* for these averaged values. We realize now that this was a mistake. These values are not unique characteristics of the different grades of printing paper because they depend upon the number and the spacing with respect to $\log SES_y$ of the paper grades used in making the prints and upon the frequency of occurrence of the DS_n values in the group of negatives printed. In fact, they have little, if any, utility for the application of the information obtained from these studies to the problem of using the density scale of a negative as an index for choosing the proper grade of printing paper to be used with that negative in order to obtain a print of excellent quality. Therefore, we are recommending now that the use of the term, *useful exposure scale*, be abandoned. No term which is entirely satisfactory for

TABLE I. First-choice prints.

Grade no.	$\log SES_y$ $(0.1\bar{G})$	Average $\log FCES$ Observed (av. DS_n)	Δ	Calculated*	Δ
00	1.78	1.85	+0.07	1.85	+0.07
0	1.35	1.38	+0.03	1.38	+0.03
1	1.24	1.29	+0.05	1.26	+0.02
2	1.11	1.13	+0.02	1.12	+0.01
3	0.94	0.90	−0.04	0.94	+0.00
4	0.54	0.54	0.00	0.50	−0.04
			0.03		0.03

* From Eq. (1): $\log FCES = 1.10 \log SES_y - 0.10$.

referring to this particular aspect of exposure scale has been found, but perhaps the phrase, *first-choice exposure scale*, is sufficiently definite and meaningful. This can be abbreviated, *FCES*, and in the logarithmic form this becomes $\log FCES$. The straight line in Fig. 2 therefore represents the relationship between the sensitometric exposure scale, $\log SES_y$, and the average of the $\log FCES$ values for the groups of first-choice prints. The equation of the least squares straight line is:

$$\log FCES = 1.10 \log SES_y - 0.10. \qquad (1)$$

In Table I the values in the third column are the directly observed average values of $\log FCES$ for the groups of first-choice prints, and in Column Δ are the corrections which must be applied to the values of $\log SES_y$ to make them agree with the observed values of $\log FCES$. Smoothed values of $\log FCES$ obtained either graphically from Fig. 2 or by use of Eq. (1) are shown in the fifth column,

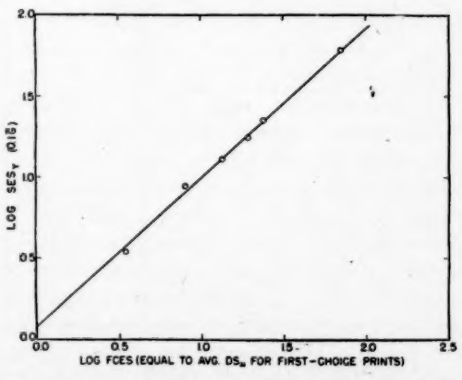


FIG. 2. Data from first-choice prints. Relation between the sensitometric exposure scales of the papers and the average of the density scales of negatives.

and the deviations between these and the values of $\log SES_y$ are shown in the last column. The fractional-gradient criterion which we have recommended for the determination of sensitometric exposure scale gives values agreeing quite closely with the observed averages applying to the groups of first-choice prints.

In order to make the information we have now obtained a useful guide for choosing a paper of known sensitometric exposure scale for the printing of a negative of known density scale, the ambiguous relationship shown in Fig. 1 must in some manner be converted into an unambiguous relationship between DS_n and some aspect of the exposure scale of the positive material. In our opinion, this is best accomplished by the psychophysical method of re-examining the prints and establishing the unique relationship which, in the opinion of a group of qualified judges, yields a group of prints having the highest possible average of print quality. It is possible, however, that if we had used a much larger number of different grades of paper in the original experiment, data could be obtained which, by the application of the methods of statistical analysis, could predict the unambiguous relationship which we desire. This we shall discuss later.

Again, we wish to emphasize that our problem assumes the use of a *few* grades of printing paper, varying with respect to exposure scale, for the printing of a large number of negatives differing over a wide range of density scale values. Let us assume, for the moment, that we have our negatives arranged in an ascending order of DS_n. It seems reasonable to assume also that the negative of lowest density scale will probably be best printed on the positive material of lowest exposure scale, or possibly on that grade having the next greater exposure scale. Hence, if we begin printing and judging the quality of results as we go, it will be found that as the density scale of the negative increases, a point will be reached where the positive material of lowest exposure scale will no longer give the best possible print that can be made from that negative. It is at this point on the negative density scale that we must begin to print on the next "softer" grade of printing paper. As the DS_n value increases further, a second point must be found where it is necessary to change from that grade to the next "softer" adjacent grade, and so on until all of the negatives have been printed, those of maximum density scale probably giving the most satisfactory prints on the positive material of greatest exposure scale. Therefore, what we must know, in order to apply this method to the printing of a large number of negatives on a few grades of positive materials, are the values of the points on the DS_n scale where the change from one grade to another must be made. These points we shall call *transition points*, abbreviated *TP*.

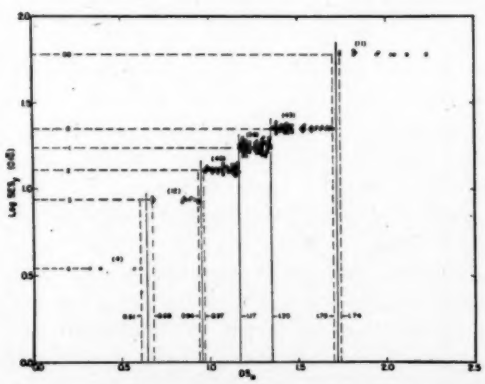


FIG. 3. Data obtained by rejudgment of the prints to determine the unambiguous relation between the sensitometric exposure scales of the papers and density scale of the negative which gives the maximum yield in print quality.

The determination of the desired transition points was made by resorting to the psychophysical method. Each one of a relatively large group of observers was requested to examine the entire group of first-choice prints, as indicated by the points in Fig. 1, and to substitute for some of them second- or third-choice prints to obtain unique groups with respect to the six grades of printing paper used. In order to do this, it was necessary to transfer prints back and forth between the various grades of paper until there was no overlap between the adjacent groups. The average judgment of the group of observers was then used in the establishment of the desired transition points.

In making the transfers which resulted in the formation of unique groups, the fundamental aim was to sacrifice as little as possible in the average photographic quality of the final group of prints as compared to the average photographic quality of the entire group (170 prints) of the first-choice

prints. In this process two subordinate principles were followed: (a) the total number of transfers from one paper grade to another should be as small as possible, and (b) the number of transfers involving an increase in the grade number should be balanced as closely as possible by an equal number of transfers to a lower grade number. Fulfillment of this second (b) condition insures that the average contrast level of the 170 prints in the final unique groups will be substantially equal to the average contrast level of the 170 first-choice prints. The values of the transition points obtained prior to the publication of our previous communication were shown graphically in Fig. 29 of that paper,[1] and these transition points relate to the unique groups as established at that time. These numerical values are as shown in the next to the last column of Table III.

Subsequent to the publication of our former paper, we discovered that the grouping shown in Fig. 29 of reference 1 did not meet condition (b) as closely as could be desired and an attempt was made to rectify this condition, keeping in mind the fundamental aim of not lowering appreciably the average photographic quality of the resultant group of 170 prints. A re-examination of the first-choice prints was therefore made by a smaller group of observers, a group selected to have an average judgment level corresponding closely to the general average of judgment level of the larger group used originally. It was found that by making slight changes in the positions of the transition points, condition (b) can be fulfilled quite closely with very little, if any, further sacrifice in print quality than was entailed by using the transition points as published in our previous communication. This new grouping is shown in Fig. 3. In the second column of Table II are shown the number of prints in the groups of first-choice prints made on the six printing materials, as indicated in the first column. In the third column are shown the number of prints in the unique groups which have now been established. It will be noted that the numbers in the various unique groups do not correspond precisely with those in the first-choice groups. The observers responsible for the final re-inspection were unwilling to reduce all of these differences to zero because this entailed the acceptance of some definitely inferior prints. According to the average opinion of the groups of observers, this would have resulted in a perceptible lowering of the average photographic quality.

TABLE II. Changes required to obtain unique groups.

Grade no.	Number of first-choice prints	Number of prints in unique groups			
		Total	First-choice	Second-choice	Third-choice
00	12	11	9	0	2
0	42	45	22	20	3
1	62	58	27	31	0
2	40	40	21	18	1
3	10	12	6	5	1
4	4	4	3	1	0
Total	170	170	88	75	7

The composition of the final unique groups with respect to the number of first-, second-, and third-choice prints is shown in the last three columns of Table II. The reader may be surprised that, in the case of the Nos. 0, 1, and 2 groups, the proportion of second- and third-choice prints is so high and may conclude therefore that the establishment of the unique groups must of necessity have materially lowered the average quality of the entire group. This, however, is not necessarily true. We wish to emphasize, as pointed out in our previous paper (see page 604 of reference 1), that in many cases the judgments were almost exactly equally divided between two prints made on adjacent grades of paper and hence the substitution of one for the other produced an extremely small decrease in the average over-all quality of the whole group.

We are satisfied now that the groupings shown in Fig. 3 and Table II represent the best average quality which is compatible with an unambiguous relation between the density scale of the negative and the exposure scale of the positive material. It will be noted in Fig. 3 that the transition points in three cases are indefinite by small amounts. This is the result of the absence of negatives having density scale values in these particular regions. Thus, the 00 group is separated from the 0 group by 0.04 on the density scale. In attempting to arrive at the most probable relationship between the transition points and the midpoints between the logSES_y values of adjacent grades of printing papers, it might be assumed that the transition point should be located at the middle of the vacant zones, but it is entirely possible that they should lie at one or the other extreme of

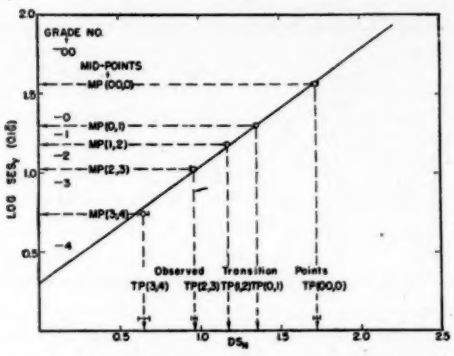


FIG. 4. Midpoints between sensitometric exposure scales of adjacent grades of paper plotted as a function of the density scales at which the transition from one grade of paper to the next was desired by the judges.

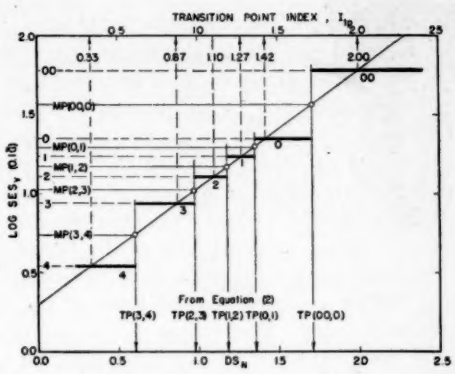


FIG. 5. Another method of representing the rejudgment data. The heavy horizontal lines show the range of DS_n values for the negatives best suited for each paper. The light lines show the relation of these ranges to: the midpoints, MP, between adjacent grades, the transition points, TP, and the transition point indices, I_{tp}.

these vacant zones. In Fig. 4 these data are represented graphically. The small circles represent the transition points, with the possible variation in three cases represented by the small vertical lines at each side of the small circles. The position of the straight line, as drawn, was determined from the small circles by the method of least squares.

The numerical data relating to Fig. 4 are shown in Table III. The transition points used for all further applications in this discussion are those given in the fifth column, and are the values read from the least squares straight line of Fig. 4, of which the equation is

$$DS_n = 1.35 \log SES_y - 0.40. \qquad (2)$$

In computing the deviations, Δ, shown in the sixth column, we have assumed that the actual transition points may lie any place between the observed limits in the cases of the first, fourth, and fifth transition points. On this basis it will be noted that the straight line represents almost perfect correlation with the observed values. The values of the transition points used in our previous communication are shown in the next to the last column. The deviations (Δ in the last column) between these points and the smoothed values shown in column 5 are very small and of little consequence from the practical point of view. However, the actual changes in the print grouping discussed previously do result in an appreciably closer approximation to the desirable principles which were followed in converting the ambiguous relationship of Fig. 1 to the unambiguous relationship of Fig. 3.

It is of interest to note how the spread of density scale values in each of the unique groups appears on the graph showing the straight-line relationship between the transition points and the midpoints between the adjacent values of logSES_y for the printing materials used. This is shown in Fig. 5, in which the spreads are represented by the heavy horizontal lines, drawn at the ordinate values corresponding to logSES_y for the positive materials. The straight line is identical in position and slope to the straight line in Fig. 4.

Suppose now that it is desired to print additional negatives using the same positive materials that were used in the determination of the straight-line relationship shown in Figs. 4 and 5. All that is necessary is to measure the density scale of a negative, locate this on the abscissa scale of Fig. 5, and then print that negative on the positive material embracing the DS_n zone in which the point falls. The values of logSES_y are unique characteristics of the positive materials since they are measured by a purely objective sensitometric method and are entirely independent of the characteristics of the negatives being printed. The transition points, as determined by our psychophysical method, are independent of the frequency of occurrence of the different values of DS_n within the group of negatives which was used in their determination. It is

therefore immaterial whether a new group of negatives to be printed has a frequency characteristic, with respect to DS_n, the same as or vastly different from that existing in the group of negatives which we used.

If a different group of positive materials is to be used, new transition points must be determined. This can be done quite simply by using the relationship represented by Eq. (2). The method followed may be either graphical or analytical, as desired. The graphical method is illustrated in Fig. 6 where the straight line is that represented by Eq. (2). The $\log SES_y$ values of four hypothetical positive materials having the grade numbers $a, b, c,$ and d (of which the $\log SES_y$ values are 1.50, 1.20, 1.00, and 0.70, respectively) are plotted on the ordinate scale at the left. The midpoints between each adjacent pair of these positive materials are determined and horizontals drawn intersecting the straight line and from these intersection points perpendiculars are dropped to the abscissa scale, thus determining transition points: $TP(a,b)$, $TP(b,c)$, and $TP(c,d)$. If additional positive materials are used covering the lower and higher $\log SES_y$ values, the same procedure is followed until all the desired transition points are located. From this point on, the use of the graphic solution is the same as that discussed in the case of Fig. 5. A transition point between adjacent members of the positive materials being used can be determined analytically from the midpoints between $\log SES_y$ values by the following equation, which is a special form of (2):

$$TP(a, a+1) = 1.35\,MP(a, a+1) - 0.40, \quad (3)$$

where

$$MP(a, a+1) = [\log SES_y(a) + \log SES_y(a+1)]/2. \quad (4)$$

To us, this method seems to be simple and straightforward and not too inconvenient, assuming that the values of $\log SES_y$ are known for the various positive materials which are to be used.

There is a slightly different method (illustrated in Fig. 6) for obtaining transition points which, in some applications, is even simpler. Along the top of this figure has been established a scale numerically identical to the ordinate scale at the left. This is designated as *transition point index*, I_{tp}. The $\log SES_y$ values of the grades of printing paper to be used are located upon the scale at the left; horizontals drawn through these points intersect the straight line represented by Eq. (2). From these intersection points, verticals are drawn to the scale of I_{tp} and a set of numerical values of I_{tp} are obtained which may be used for computing directly the desired transition point, TP. The latter computation is a simple mathematical operation, since it is only necessary to take the average of any two adjacent values determined from the I_{tp} scale:

$$TP(a, b) = [I_{tp}\,(\text{for } a) + I_{tp}\,(\text{for } b)]/2. \quad (5)$$

It may be interesting now to go back and determine, by the method described in the construction of Fig. 6, the transition point indices for the six papers used here. On referring now to Fig. 5, it will be seen that a scale of transition

TABLE III. Revised transition points. Correlation with midpoints between values of $\log SES_y$.

Grade no.	$\log SES_y$	Mid-points	Transition points (DS_n) Observed	Smoothed*	Δ	Former values	Δ
00	1.78						
		1.57	1.70 to 1.74	1.71	0.00	1.70	+0.01
0	1.35						
		1.30	1.35	1.35	0.00	1.33	+0.02
1	1.24						
		1.18	1.17	1.19	+0.02	1.18	+0.01
2	1.11						
		1.03	0.94 to 0.97	0.98	+0.02	0.94	+0.04
3	0.94						
		0.74	0.61 to 0.68	0.60	−0.01	0.65	−0.05
4	0.54						

* From Eq. (3).

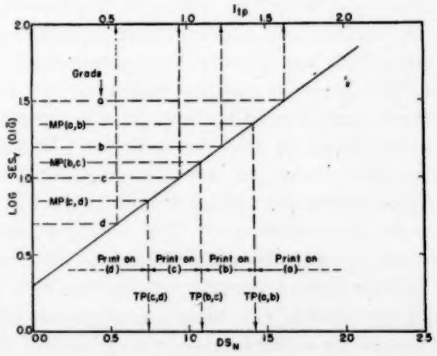


FIG. 6. Method of determining transition point indices, I_{tp}, for four hypothetical papers (a, b, c, and d) and the use of these indices in determining transition points.

TABLE IV.

LogSES_y	Transition point (DS_n)	Transition point index (I_{tp})	$I_{tp}-\log SES_y$
1.78		2.00	+0.22
	1.71		
1.35		1.42	+0.07
	1.35		
1.24		1.27	+0.03
	1.19		
1.11		1.10	−0.01
	0.98		
0.94		0.87	−0.07
	0.60		
0.54		0.33	−0.21

TABLE V.

Grade no.	Values from the unique groups: Average DS* Observed	Smoothed	Mid-points	Transition points	LogSES_y	Δ
00	1.94	1.96			1.78	+0.18
			1.68	1.71		
0	1.48	1.41			1.35	+0.06
			1.33	1.35		
1	1.26	1.26			1.24	+0.02
			1.18	1.18		
2	1.09	1.10			1.10	−0.01
			0.98	0.97		
3	0.84	0.88			0.94	−0.06
			0.62	0.60		
4	0.47	0.37			0.54	−0.17

* Average DS_n is equal to the average logES of the prints in the unique group (Fig. 3).

point indices has been established along the upper boundary. Perpendiculars drawn from the points where the heavy horizontal lines intersect the straight line represented by Eq. (2) give the desired values of the transition point indices. The numerical values are shown in the third column of Table IV. The relationship between logSES_y (first column) and the transition point index for our six positive materials may now be seen, the differences between these two values being shown in the last column of Table IV.

These values of I_{tp} are unique characteristics of each positive material. They depend for their magnitude only upon the relationship represented by the straight line in Fig. 6. They are not dependent upon the number and spacing of the positive materials which compose the new group of positive materials to be used in printing, nor are they dependent upon the frequency distribution of density scale values within the group of negatives to be printed. They are, however, dependent upon the general level of print contrast which is represented by Eq. (2). We have not thus far emphasized this fact, but it should be recognized that Eq. (2) actually depends upon the general level of print contrast which our group of observers considers to be best possible quality. We know that, on the average, the 170 first-choice prints have an average contrast level somewhat below that which might be preferred by some photo-finishers in the amateur field. Therefore, it must be remembered at all times that prints made according to our findings, as exemplified by Eq. (2), have a general contrast level somewhat softer (and, in the opinion of our thirty observers, of better photographic quality) than that which is found in some amateur fields.

Aside from this, the values of I_{tp} are unique characteristics of the positive materials to which they apply. An I_{tp} value may be identified as a particular aspect of the log exposure scale of the positive material, logES_y, which we might be tempted to refer to as the *useful exposure scale.* Such a designation, however, is confusing and is, in fact, misleading in its implication. Thus, for our 00 grade, I_{tp} is found to be 2.0, but by consulting reference 1 (Fig. 20, p. 592) it will be seen that logES measured from the tangent point s (see reference 1) to the point of zero gradient on the extreme toe is only 1.8. It is undesirable to assign a UES value greater than the maximum available logES_y. The only "useful" aspect of I_{tp} is for finding the transition point, in terms of DS_n, between two adjacent grades of positive material. Thus far at least, we have not found any numerical value of logES_y to which the term *useful* can be applied in the sense originally attributed to the phrase *useful exposure scale.* The connotation which has been attached to this term implies that it is directly indicative of that portion of the $D-\log E$ characteristic curve of a positive material which is used in making an excellent print from a single negative or, more generally, indicative of the average DS_n value of a group of negatives, all of which will yield prints of excellent quality when printed on that particular grade of positive material. It has become evident now that there is no unique value of logES_y to which can be attributed the qualities implied by the expression *useful exposure scale.*

Thus, it is now apparent that the average value

of DS_n for a group of negatives yielding prints of satisfactory quality on a given positive material is dependent upon: (a) the number of grades used in a given case and their spacing with respect to each other on the log exposure scale; and (b), the frequency of occurrence of the DS_n values within the group of negatives being considered. Therefore, we are advocating the complete abandonment of the term *useful exposure scale* since, as we now see the problem, it can serve no useful purpose and can only lead to unwarranted conclusions. Values of I_{tp} are not indicative, nor are they intended to be indicative, of the average density scale of the negatives or of the range of DS_n values which, according to Figs. 5 and 6, must be printed on a specified grade of positive material. They are intended only for the determination of the transition points between adjacent grades of paper. For this reason we shall *not* refer to this particular aspect of exposure scale as *useful exposure scale*. The term *conditional exposure scale* ($\log CES_y$) might be used since the numerical value of $\log CES$ for a given positive material would be dependent upon the one *condition* as mentioned above, namely, the general contrast level which is considered to be most desirable. However, the term *transition point index* seems preferable since it is not easily confused with *sensitometric exposure scale* and it suggests more clearly the purpose for which the values are to be used.

In our previous communication[1] we did not make any use of the average values of DS_n for the unique groups as shown in Fig. 3 (Fig. 29 of reference 1), nor, in fact, did we publish the values of these averages in that paper. The question may be raised as to why we did not use these values as possible indications of the *useful exposure scale*. Perhaps it would be well to consider these average values and their relation to the $\log SES_y$ of the positive materials. The average values of the density scale, av. DS_n, of the negatives used in making each of the revised unique groups of Fig. 3 are given in the second column of Table V. The relationship between these values and those of $\log SES_y$ for the positive materials is shown in Fig. 7. The solid straight line, of which the equation is

$$\text{av. } DS_n = 1.28 \log SES_y - 0.32, \qquad (6)$$

was determined by the method of least squares. It will be noted that this straight line does not fit the experimentally determined points very well and there is some question as to whether or not the relationship between these two variables should be linear. However, assuming, for the time being at least, that the relationship is satisfactorily represented by the straight line, the smoothed values of average DS_n are as shown in the third column of Table V. These values, of course, could also be identified with an aspect of the log exposure scale of the positive materials. The midpoints between adjacent values of average DS_n are shown in the fourth column and they may be compared with our smoothed psychophysically determined values of transition points derived from Fig. 4 which are shown in the fifth column. The values in the fourth column agree well with those in the fifth column. This agreement is to be expected since the transition points were used to establish the limits for the unique groups from which the average DS_n values were derived. To attempt to use the midpoints between these average DS_n values as a means of deriving transition points is futile since the answer must be known in advance. The procedure is effectively a "vicious circle."

It may be worth while to consider the possibility of applying to our experimental data the methods of statistical mathematics as discussed by Sanders.[3] Our experimental data may not be satisfactory for this purpose because we used only a relatively few grades of positive materials. Sanders' treatment is based fundamentally on

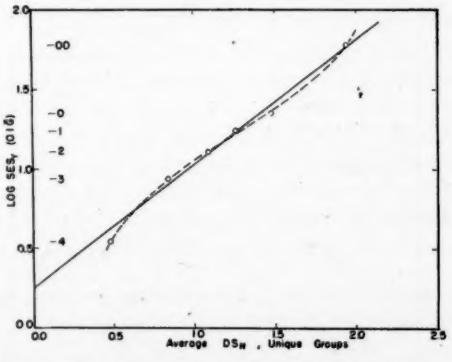


FIG. 7. Relation between the sensitometric exposure scales of the papers and the average of the density scales of the negatives for the unique groups of Fig. 3.

the assumption that experimental data similar to ours, as shown in Fig. 1, have been obtained by using a large number of negatives (which we did) and also a large number of positive materials closely spaced with respect to log exposure scale. This latter condition our experimental work does not fulfill. Our data in Fig. 1 appear to be adequate for the determination of the line of regression of x on y, which is a part of Sanders' mathematical treatment. In fact, line A in Fig. 2 is equivalent to the regression line x on y as defined by Sanders. Difficulties are encountered, however, when we attempt to determine the line of regression of y on x. This results because of the relatively few positive materials which we used. There are not a sufficient number of $\log SES_y$ values for each value of DS_n to obtain a reliable average value of $\log ES_y$ to enable us to plot with certainty the line of regression of y on x. However, we can at least make an attempt, and as a first approximation a series of equal intervals, $\Delta DS_n = 0.10$, were established and the average of all of the $\log SES_y$ values falling within each of these intervals was determined. These average values were then plotted at the midpoint of each of the intervals. The result is shown in Fig. 8, the position of the solid straight line having been established by the method of least squares. The equation of this line is

$$DS_n = 1.60 \log SES_y - 0.69. \qquad (7)$$

Now, if this regression line is used in accordance with Sanders' treatment, values of transition points are obtained as shown in the second column of Table VI. The disagreement between these values and our psychophysically determined transition points, shown in the fourth column, is so great as to make it obvious that the statistical method applied in this manner to our data does not give the desired result.

However, the approximation method employed in the construction of the line of Fig. 8 is subject to some criticism in that it tends to give much more weight to very low and high values of DS_n than to the intermediate ones which depend upon a much larger number of observations.

There is a more laborious, though perhaps more rigorous, method of computing regression lines from a scatter diagram such as that shown in Fig. 1. This method consists in fitting the lines to the original data points, using the method of least squares with no preliminary averaging of the data. The regression line of y on x is obtained by making the squares of the deviations from the line in the y direction a minimum, and the regression line of x on y is obtained by making the squares of the deviations in the x direction a minimum. This method has been applied to Fig. 1 and the resulting two regression lines are shown in Fig. 9. The equation for the y on x regression line is:

$$DS_n = 1.75 \log SES_y - 0.90. \qquad (8)$$

This equation, as well as Eq. (6), differs considerably from Eq. (2) which, it may be recalled, was based on the opinion of the judges regarding the transition points which give the highest yield of high quality prints.

The transition points derived from Eq. (7), or from the y on x regression line of Fig. 9, are

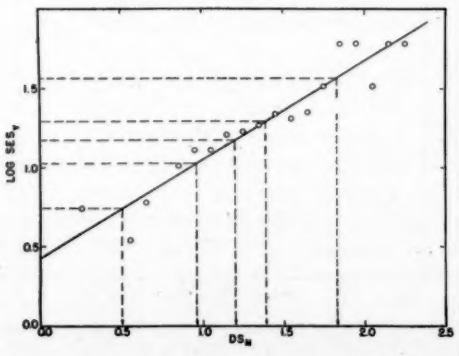


FIG. 8. Computation of the line of regression of y on x.

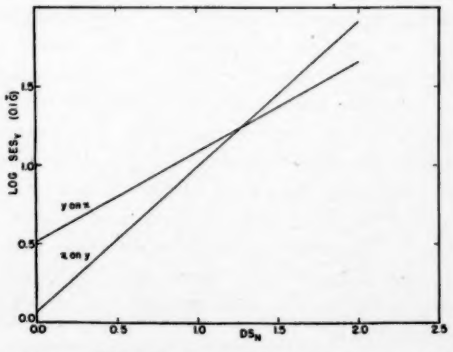


FIG. 9. Regression lines y on x and x on y, computed by the method of least squares.

shown in the third column of Table VI. It can be seen that the agreement with the observed transition points is not good for the high and low values.

We conclude that the statistical method of computing transition points from the regression line of y on x is not applicable *to our data* since it does not give results in good agreement with the transition points obtained by direct judgment of the prints. It should be kept in mind that the use of the regression line of y and x, as proposed by Sanders, is based on the assumption that the frequency distribution of the data points in vertical cross sections taken through the scatter diagram is symmetrical and that the number of grades of paper used, as well as the number of negatives, is large. It is obvious from an inspection of Fig. 1 that our data do not meet these conditions. Our data were obtained by the printing of a large number of negatives on a relatively *few* grades of paper. Consequently, it perhaps should not be expected that the mathematical and the psychophysical methods of approach should yield identical results. We have no objection to the use of a large number of different grades of paper for printing each of the large number of different negatives as a means of attacking this problem but we wish to point out that it would be an enormous task considering that all of the prints have to be judged by a number of judges. As far as we know, no one has used this method of approach and hence there are no data available for evaluating this procedure.

While the mathematical treatment involving the regression line of y on x, as proposed by Sanders, does not yield results in quantitative agreement with our experimental results, it does, nevertheless, predict in a qualitative way the same general principle which we found to be true from the second part of our psychophysical judgment data, namely, that for the soft papers the density scale of the negatives tends to exceed the sensitometric log exposure scale of the paper, and for the hard papers the density scale of the negatives tends to be less than the sensitometric log exposure scale of the paper. Thus, the mathematical and the experimental results support each other as to the general nature of the result but differ in degree. We propose, in the application of our results to the practical problem of selecting the grade of paper for printing any negative, to recommend the quantitative relationship given by our experimental data (Fig. 5 and Eq. (2)). The mathematical results are unnecessary as far as the practical solution of the problem is concerned, but they are of interest when an attempt is made to find a theoretical explanation of *why* it is that the density scale of the negative should sometimes be greater than and sometimes less than the sensitometric exposure scale of the paper.

TABLE VI.

Grade no.	Transition points: Computed from regression line y on x, Method a	Transition points: Computed from regression line y on x, Method b	Transition points, Observed	Transition point indices: Computed from regression line y on x, Method a	Transition point indices: Computed from regression line y on x, Method b	Transition point indices, Observed
00				2.17	2.23	1.97
	1.83	1.85	1.70			
0				1.48	1.47	1.40
	1.44	1.37	1.35			
1				1.31	1.28	1.25
	1.20	1.16	1.17			
2				1.09	1.04	1.08
	0.96	0.90	0.87			
3				0.82	0.74	0.85
	0.50	0.39	0.60			
4				0.18	0.05	0.33

It may be well to repeat that in some cases, in the application of our data to practical printing, it may be advantageous to assign to each paper a *transition point index*, I_{tp}, which is obtained from the sensitometric exposure scale by the use of Fig. 5 or an equation obtained from this figure. These transition point indices have, in certain cases, the advantage that it is unnecessary for the operator who makes the prints to use an equation or a graph, since the transition points which he needs in order to decide what negative to print on each grade of paper are simply the midpoints between the transition point indices for adjacent grades of paper. This midpoint is, of course, given by the simple arithmetic average of two transition point indices. This part of the procedure is essentially equivalent to that suggested by Sanders,[3] although the terminology is different.

III. SENSITOMETRIC EXPOSURE SCALE AS A BASIS FOR THE SPECIFICATION OF THE GRADES OF PHOTOGRAPHIC PAPER

We suggest that the sensitometric exposure scale[2] be adopted as the basis for the specification

of the grades of photographic papers. A careful study has been made in these Laboratories of the requirements which should be met by such a specification primarily from the standpoint of the user of these materials. The following conclusions have been reached:

1. The *sensitometric exposure scale* is the most suitable basis for the derivation of grade numbers because it bears a significant and useful relation to the density scales of the negatives which in practice give the best prints on each grade of paper.

2. The *contrast* of the paper is not satisfactory as a basis for deriving grade numbers. For papers differing in $D-\log E$ curve shape or in density scale, it gives equal ratings to papers which are unsuitable for printing the same negatives and unequal ratings to papers which are suitable for printing the same negatives. The term *contrast* relates to the subjective appearance of prints made on the paper. While the sensitometric exposure scale is an important factor in determining the contrast of the paper, other essential factors are the *density scale* of the paper and the *shape* or *gradient* characteristics of the $D-\log E$ curve. If these other factors are included in the derivation of a grade number, the number will no longer be indicative of the density scales of the negatives which give the best prints, and the practical significance and utility of the grade numbers will be seriously impaired.

3. The term *scale index* is suggested as a means of designating the grade of paper. The present grade numbers, letters, or verbal terms now assigned by the paper manufacturer need not then be abandoned abruptly, since both the new and the old designations could continue to exist without confusion. The old terms could, in fact, be continued as long as the manufacturer wishes, simply as an identification of his particular papers, but the scale index values for each grade would, of course, also be given. The terms *sensitometric exposure scale* or *log exposure scale* are too long for general use.

4. The numerical values for *scale index* should be large enough so that they cannot easily be confused with the present grade numbers. The present numbers vary from 0 to 5, whereas the new values might all be greater than 10. For example, scale index might be 100 times the sensitometric exposure scale ($100 \log SES_w$). This has the advantage that scale index is obtained from $\log SES_w$ simply by removing the decimal point. In general, scale indices would then lie between 50 and 200 for present papers. Thus, for a paper made in only three grades, the following values might, for example, occur:

Grade	Sensitometric exposure scale (logarithmic)	Scale index
1	1.40	140
2	1.10	110
3	0.85	85

For a paper made in six grades, the following values might occur:

Grade	Sensitometric exposure scale (logarithmic)	Scale index
0	1.55	155
1	1.30	130
2	1.10	110
3	0.90	90
4	0.70	70
5	0.55	55

If it is desired that the numbers lie between 25 and 100, scale index might be defined as 50 times $\log SES$.

5. It is not practical to require a definite spacing of the various grades. A paper made in only two or three grades, for example, is usually intended for use in a different type of photographic work than a paper made in six grades. The spacing which meets the practical requirements of one type of work is not, in general, the best for another. Consequently, the spacing of the grades is not a suitable subject for standardization. Each grade should be assigned a scale index, but no restrictions should be placed on what the values should be.

It may be well to review briefly the experimental findings which have shown why the *contrast* of the paper is not a satisfactory basis for assigning grade numbers. The term *contrast* refers to the *appearance* of the prints. It may have been assumed by some that the present grade numbers or designations assigned by manufacturers are based on the *contrast* of the paper because the term *contrast grade number* is sometimes used. The fact is, however, that the present grade

numbers are more indicative of exposure scale (which is one component of contrast) than of contrast itself.

The numerical evaluation of the *contrast* of photographic papers has been studied for a number of years. The results of an extensive investigation of this subject in these Laboratories were published[5] in 1926 and 1927. It was concluded that contrast (Ω) is dependent upon two factors, as follows:

$$\Omega = f\,(\text{rate, extent}). \qquad (9)$$

The *rate* factor was evaluated by means of the average gradient, $\bar{G}(D)$, of the utilized portion of the $D-\log E$ curve of the paper determined with respect to *equal increments in density*, D. It must be kept in mind that G is a gradient and is equal to $dD/d\log E$. The fundamental reason for evaluating the average gradient of the characteristic curve between the two limiting points in terms of equal increments of density rather than in terms of equal increments of log exposure, $\log E$, is that the eye in viewing a print perceives density differences, not differences in log exposure. The value of $\bar{G}(D)$ is dependent upon the *shape* of the curve between two limiting points.

The *extent* factor in the contrast equation was evaluated by means of the utilized density scale, $D_{\max}-D_{\min}$. Thus, contrast was determined numerically by the formula

$$\Omega = k\bar{G}(D)\times(D_{\max}-D_{\min}), \qquad (10)$$

where k is a constant which was related inversely to the square of the *just-noticeable density difference* which the eye can see. In terms of this unit, k is 20,000, but for practical work it was suggested that a constant of 20 be used to make the contrast numbers convenient in size.

Equation (10) has been in use for a number of years in these Laboratories and considerable evidence has accumulated which demonstrates its validity. In fact, the data appear to indicate that it is superior to any other formula which has thus far been proposed for evaluating contrast. The use of the average gradient, $\bar{G}\,(\log E)$, of the utilized portion of the $D-\log E$ curve is sometimes suggested as a criterion of contrast, but this is equivalent to using the slope of the straight line drawn between the two selected end points of the curve. Consequently, this criterion ignores differences in the shapes of the curves for different papers and, as shown by practical tests, does not give contrast values which correlate with the appearance of prints made on the papers.

The slope (γ) of the straight-line portion of the $D-\log E$ curve has also been proposed as a measure of contrast, but it has the defect that it ignores differences in maximum density and in the shape of the toe and shoulder. For most papers, the straight-line portion represents only a small part of the curve and hence its slope (γ) cannot serve as a significant index of the entire curve shape.

In 1939 a communication[6] from these Laboratories on the sensitometry of photographic papers presented further evidence regarding the correlation of the contrast values obtained by Eq. (10) and the observed contrasts of the various grades of several types of papers then on the market. This evidence indicated that differences in the subjective aspect of contrast were more directly related to logarithmic than to arithmetic differences in Ω. Consequently, the use of log Ω as a practical index of contrast was suggested. The Ω formula ranks papers in the proper *order* of contrast, but log Ω gives a better evaluation of the magnitude of the *contrast differences* between them.

The purpose of the present discussion is to emphasize that it is not the lack of a significant method for evaluating contrast which leads us to recommend the sensitometric exposure scale as the basis for grading photographic papers. Contrast can be evaluated satisfactorily in quantitative terms by the Ω formula, but, nevertheless, sensitometric exposure scale is preferable as a basis for specifying grade numbers. The reason for this is that papers which have the same contrast number may have widely different curve shapes and widely different exposure scales. Because of the difference in exposure scales, such papers are unsuitable for printing the same negatives, as will be demonstrated in connection with Figs. 10 and 11. From the standpoint of the consumer, therefore, it seems desirable to grade papers in terms of exposure scale since the values

[5] L. A. Jones, "The contrast of photographic papers," J. Frank. Inst. **202**, 177, 469, 589 (1926); **203**, 111 (1927); **204**, 41 (1927).

[6] L. A. Jones and C. A. Morrison, "Sensitometry of photographic papers," J. Franklin Inst. **228**, 445 (1939).

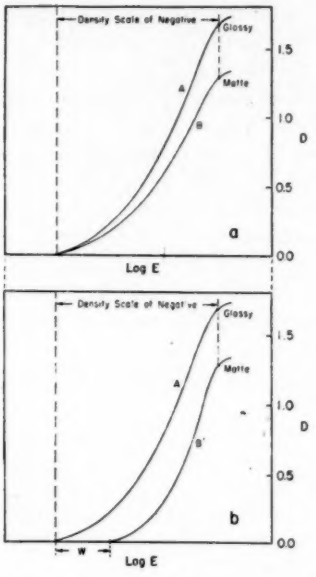


FIG. 10. Papers *A* and *B* in the upper graph have the same exposure scales and print the same negatives successfully and, hence, should be assigned the same grade designation even though they differ in contrast. Papers *A* and *B'* in the lower graph have different exposure scales and do not print the same negatives successfully and, hence, should not be assigned the same grade designation even though they may be assigned equal contrast ratings.

will then be more indicative of the negatives which are suitable for printing on each paper.

For a family of photographic papers which have very nearly the same maximum density and homologous*** $D-\log E$ curve shapes but which differ progressively in exposure scale, a set of grade numbers derived from exposure scale would be consistent with a set of grade numbers derived from *contrast.* It is in comparing papers having different maximum densities (usually due to different surfaces, such as glossy, semimatte, matte, etc.) or having appreciably different $D-\log E$ curve shapes, that the relation between grade and contrast breaks down.

Consider first the case where the papers differ widely in maximum density as shown in Fig. 10a. These curves are typical of the difference between glossy and matte surface papers having equal exposure scales. Careful tests have shown that these papers print the same negatives successfully. In Fig. 10b are shown curves for a glossy paper and a matte paper which would be rated alike in terms of contrast. Tests show that these papers do not print the same negatives successfully. In practice, the user may, for instance, have made a print on the glossy paper in Fig. 10a and decided that it was satisfactory in exposure and in exposure scale but that the subject matter was, for aesthetic reasons, especially suited to a matte surface paper. He should, we believe, be able to choose, for making another print, a matte surface paper having the same grade number as the glossy paper and obtain a satisfactory print. This will be the case if the *sensitometric exposure scale* is used as the basis for grading but it will not be the case if the *contrast* of the paper is used as the basis. To obtain a matte paper having a *contrast* equal to that of the glossy paper, it is necessary to choose a grade of matte paper having a shorter exposure scale, as shown in Fig. 10b. This matte paper must have a higher gradient and a shorter exposure scale to make up for the fact that its maximum density is less than that of the glossy paper. (It is unnecessary for this illustration to show *exactly* how the matte paper should be altered in gradient and exposure scale to match the glossy paper in *contrast;* the change represented by these two figures is approximately correct and is typical of the change which would be made simply by choosing a different grade of matte paper.) If these two curves in Fig. 10b were assigned the same grade number, the user of the paper would be misled, for the matte paper, *B'*, is wholly unsuited for printing the negative which was satisfactory on the glossy paper, *A*. Its exposure scale is too short for the density scale of the negative. Much of the highlight detail and some of the middle-tone detail are reproduced as blank white paper; or, if the print exposure is increased to improve the highlights, the shadows are impaired by being pushed onto the shoulder of the curve. Tests have shown conclusively that the average user prefers the matte paper and glossy papers having the same exposure scale (and hence the same scale index) when the same negatives are to be printed on each. In fact, he expects and in many cases desires the prints on matte paper to have lower contrast than the prints on glossy paper. Consequently, in this case it is clearly desirable to

*** The curves are homologous in shape if the abscissa values for all of the curves are proportional to the corresponding abscissa values for any one of the curves.

assign grade numbers on the basis of *exposure scale* rather than on the basis of *contrast.*

Consider next the case in which the papers differ widely in curve shape but have the same maximum density. This is illustrated in Figs. 11a and 11b. Papers A and A' belong to a family of papers differing in exposure scale but having the same type of curve shape. Paper B belongs to a different family of papers having curves which are considerably more bowed. In Fig. 11a the two grades chosen have the same sensitometric exposure scale, and we believe it is desirable to assign them the same grade number even though the paper represented by Curve B is generally considered to have higher *contrast.* These two papers accommodate the same negative density scales, and this is the condition which is most important from the standpoint of the user. It may be argued that paper B should be assigned a higher grade number than paper A, since B has higher contrast. This, however, leads to the serious practical difficulty shown in Fig. 11b. Curve A' has the same general shape as Curve A of Fig. 11a, but A' has a shorter exposure scale. If contrast were adopted as the basis for assigning grade numbers, papers A' and B would be assigned the *same* grade number. The difficulty is that these two papers are unsuitable for printing the same negatives. Paper B gives its best prints from negatives which have longer density scales than those which would print best on paper A'. Paper A' has too short an exposure scale to accommodate the negatives which are most satisfactorily printed on paper B. For such negatives, the highlights are not reproduced at all on paper A'. If the print exposure is increased, the highlights are improved but the shadows are ruined by being pushed onto the shoulder. Thus, it can be seen that a contrast grading system which would assign different grade numbers to the papers in Fig. 11a would make the serious error of assigning the same grade numbers to the papers in Fig. 11b. The user of the paper would be led to believe that the papers in Fig. 11b could be used interchangeably for printing the same negatives, whereas practical tests show that papers having the same exposure scale are more suitable for this purpose.

It is important to keep in mind that the use of the sensitometric exposure scale as the basis for assigning grade numbers does not insure that *identical* or *matched* prints will be obtained from the same negative on two different papers which are assigned the same grade number. It is to be expected that the prints may differ considerably in *appearance,* but they will be the best prints which can be made in the sense that *less desirable* prints would result from using any of the other grades of the *same type* of paper. The use of contrast as a basis for grading does not give a better solution to the problem; it introduces the more serious practical difficulty that papers assigned the same numbers may be wholly unsuited for printing the same negatives.

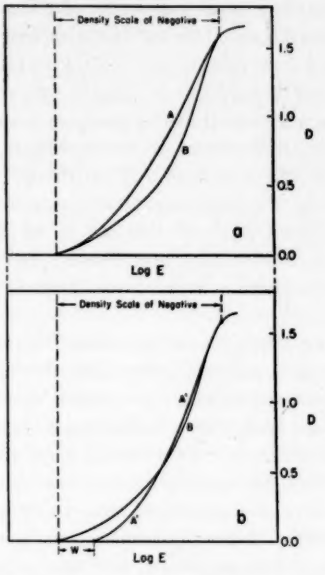


FIG. 11. Papers A and B in the upper graph have the same exposure scales and print the same negatives successfully and, hence, should be assigned the same grade designation even though they differ in contrast and curve shape. Papers A' and B in the lower graph have different exposure scales and do not print the same negatives successfully and, hence, should not be assigned the same grade designations even though they may be assigned equal contrast ratings.

The use of gamma (γ), the slope of the straight-line portion of the $D-\log E$ curve, or various approximations of gamma are sometimes suggested as a basis for grading papers, but, like the use of contrast, they rate papers alike which are wholly unsuited for printing the same negatives. Since the straight-line portion of some papers is very short, gamma often ignores the major part of the

curve and is a poor indication of either contrast or exposure scale. The various approximations of gamma do not adequately take into account the toe of the $D-\log E$ curve. This is a defect because experiments show that the eye is very sensitive to luminance differences in the highlight region of the print which is rendered on the toe and that changes in toe shape are readily apparent in the prints. The length of the toe is an important factor in determining the density scales of the negatives which will give the best prints on a paper.

Many of the criteria proposed for the grading of papers have been attempts to obtain a simple method of measuring the *contrast* of the paper. This arises from the erroneous assumption that contrast is the characteristic of a printing paper which is of maximum importance to the consumer. It is now apparent that the consumer is interested in obtaining papers which will satisfactorily fit his negatives. For this purpose, the sensitometric exposure scale, and not contrast, is of maximum value.

It is *impossible* to obtain a *single* number which is an adequate indication of both *exposure scale* and *curve shape*. Of prime importance is the specification of the grade of paper in terms of exposure scale. Therefore, we suggest the adoption of the *scale index* for the grading of photographic papers.

IV. CURVE SHAPE

As a subsequent step it may be desirable to decide upon a method of specifying curve shape in numerical terms. Exposure scale and density scale having been defined, a specification of curve shape which is adequate for practical purposes can be obtained by including a measurement of the *amount* and the *nature* of the bowing or curvature of the $D-\log E$ curve, as proposed by Morrison.[7]

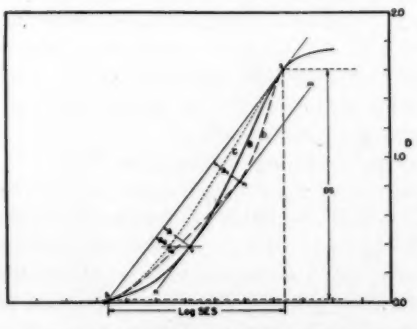


FIG. 12. Specification of the D-log$E$ curve shape of three papers: C, B, and D.

The problem of specifying curve shape is of sufficient importance to deserve some further discussion at this time. It may be that Morrison's treatment of this subject has escaped the attention of many workers in this field or at least has not yet received the serious consideration which it deserves. The method is illustrated in Fig. 12. The heavy curve designated as B represents the $D-\log E$ characteristic of a typical developing-out paper. The points, h and s, are located according to the method previously discussed as defining the extent and position of the sensitometric exposure scale, logSES, the gradient, $dD/d\log E$, at the point, h, being one-tenth of the gradient, $\bar{G}$, of the straight line, hs. The straight line, nm, is drawn parallel to hs and tangent to curve B at the point designated as v. The gradient at this point is $\bar{G}$. The positions of the three points, h, v, and s, together with the gradients at these three points, have been found to be a satisfactory and unambiguous specification of the position and shape of the curve, B. The position of the point, v, relative to that of h and s is a satisfactory index of the amount and nature of the bowing of the curve between the points, h and s. The *amount* of this bowing can be conveniently expressed in terms of the length of the line drawn from v normal to the line, hs, and this quantity is termed *arcuance*, A. The *magnitude* of arcuance is measured in terms of the scale units used in the construction of the sensitometric curve and, in order to make the number convenient for use, it is customary to multiply the actual measured distance in scale units by 100.

It is evident that a large number of curves, differing in shape from that of B, can be drawn through the points, h and s, all having the same gradient characteristics at the points, h, v, and s. One group of such curves may have the same value of arcuance as curve B but differ among themselves with respect to the position of the vertex. One such curve, D, is shown in Fig. 12. In

[7] C. A. Morrison, "An objective method of specification for sensitometric curves of photographic papers," J. Franklin Inst. **243**, 55 (1947).

this case, v_1, the vertex of the curvature, lies at a density value considerably higher than the vertex, v, of curve B. The position of the vertex can be expressed conveniently in terms of *vertex ratio*, V, which is defined by the relationship

$$V=[(D_v-D_h)/(D_s-D_h)]\cdot 100.$$

Thus, the vertex ratio is the magnitude of the density difference between points v and h expressed as a percentage of the density scale, DS, of the material. It is also evident that a number of curves differing in shape from that of B can be drawn through the points, h and s, having values of arcuance differing from that of B with constant value of vertex ratio. Such a curve is illustrated in Fig. 12 by the dotted curve, C. In this case the vertex lies at the point, v_2. The arcuance value, A_2, is in this case one-half of the arcuance value for the curve, B, while its vertex ratio is the same as that for B.

Morrison has shown that the characteristic curve of any of a large number of commercially available developing-out papers can be reconstructed satisfactorily, provided the coordinate values of the three points, h, v, and s, are known. The coordinate values of these three points, as established by the procedure just discussed, therefore become satisfactory and meaningful specifications of *curve shape* as well as of other useful sensitometric characteristics of the material in question. Thus, the coordinates of these three points may be written:

For point s: $x=\log E_s$, $y=D_s$.
For point v: $x=\log E_v$, $y=D_v$.
For point h: $x=\log E_h$, $y=D_h$.

In addition to the *positions* of these three points, the gradients at these points are also specified by virtue of the method of establishing the positions of the three points. Thus, the gradients, G_s and G_v, at the points, s and v, are equal to $\bar{G}$, the average gradient of the straight line, hs, while the gradient, G_h, at the point, h, is equal to $0.1\bar{G}$. To emphasize the simplicity and the completeness with which the coordinates of the three critical points, h, v, and s, determine all the useful sensitometric characteristics of a particular printing paper, the following list is given:

Log exposure scale: $\log ES=\log E_s-\log E_h$.
Density scale: $DS=D_s-D_h$.
Maximum useful density: $D_{max}=D_s$.
Average gradient: $\bar{G}=DS/\log ES$.
Speed: $S=1/E_s\cdot$a constant, c.
Vertex ratio: $V=100\cdot[(D_v-D_h)/DS]$.
Arcuance: $A=$ the length of the straight line drawn from the vertex of the curve, v, normal to the line, hs. This distance is most conveniently expressed as 100 times the length of this line measured in the coordinate scale units.

Morrison, in his discussion of this subject, has described an instrument termed a "curve duplicator" (see Fig. 2 of reference 7). By using the values of the coordinates of the three points, h, v, and s, this instrument can be adjusted so that a very close duplication of the curve from which the coordinate values were read can be obtained. The performance of this instrument has been tested on a large number of different printing papers of all types, grades, and surface finishes made by several manufacturers.

In our opinion it has been demonstrated beyond doubt that the coordinates of the three critical points, h, v, and s, supplemented by the gradients at these three points, are significant and unambiguous specifications of curve shape and are, for all practical purposes, entirely satisfactory. We wish to point out that the positions of these three critical points have not been chosen arbitrarily. The positions of the points, h and s, were established by our own work and are derived from the characteristics of the best possible prints which we were able to make from 170 different negatives, using a group of printing papers consisting of six grades ranging with respect to sensitometric log exposure scale from 0.54 to 1.78. The vertex point, v, was chosen by Morrison as being the point which most significantly represents the extent and the type of bowing between the end points, h and s. That this vertex point fulfills this expectation is amply verified by the subsequent success with which the coordinates of the three points permitted the reconstruction of the shape of the curve from which these coordinate points were derived.

In further support of the desirability of adopting these three significant points for the specification of curve shape, it should be pointed out that sensitometric characteristics derived therefrom are useful in the prediction of the quality of objective tone reproduction obtainable with certain combinations of negative and positive ma-

terials. This is illustrated in Fig. 13. In the left-hand section of this figure is shown a characteristic curve, P, for a typical printing paper. The ordinates indicate values of log exposure, $\log E_y$, while the abscissa values are for density, D_y; curve L in the right-hand section of the figure represents the characteristic curve of a *negative*, not of a negative material. Thus, it represents the functional relationship between negative density, D_n (the ordinates), and the log luminance of the object, $\log B_o$.

$$D_n = f(\log B_o). \tag{11}$$

For perfect objective reproduction, curve L must be the mirror image of curve P, which is merely a different way of saying that for any pair of corresponding points, such as f and g, the following relationship must hold:

$$dD_g/d\log B_o \cdot dD_f/d\log E_y = 1. \tag{12}$$

The arcuance, A_y, of the positive material must then be equal to the arcuance, A_n, of the negative, while the vertex distance, V_y, for the positive material must be equal to the vertex distance, V_n, for the negative, where V_n is defined by the relationship

$$V_n = [\log B_o\,(\text{at } v_n) - \log B_o\,(\text{at } h_n)]/\log BS_o. \tag{13}$$

Similar significant relationships can be established between the characteristics of the printing material expressed in terms of coordinates of the three critical points and the characteristics of other negatives, one of which is illustrated by curve O in the right-hand section of Fig. 13. In this case it is impossible to achieve perfect objective tone reproduction because the brightness scale of the scene represented by negative O is greater than the density scale of the positive material. Proportional reproduction can be obtained, however, and curve O, as plotted, is such as to give this. With proportional reproduction, as with perfect objective reproduction, the vertex ratio of the positive material is equal to the vertex ratio of the negative. The arcuance of the positive material, however, is then related to the arcuance of the negative as follows:

$$A_y = A_n(\cos\theta_n/\sin\theta_y), \tag{14}$$

where θ_n is the angle between the average gradient line for the negative and its $\log B_o$ axis, and θ_y is the angle between the average gradient line for the positive material and its $\log E$ axis. A comparison of arcuance and vertex ratio values can therefore give valuable information as to the quality of the objective tone reproduction.

Considerable space has been devoted to the discussion of this method for the specification of the curve shape of positive materials because we are convinced that it is the best proposal that has thus far been advanced for this purpose. We wish to urge strongly, if it seems desirable to

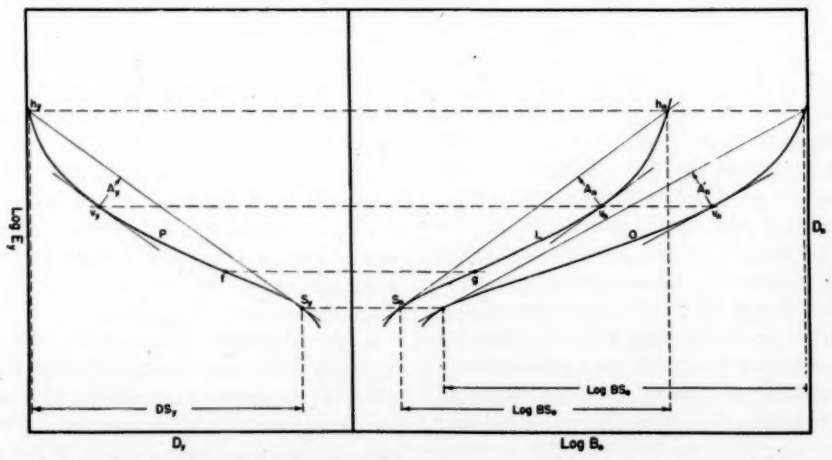


FIG. 13. Specification of the curve shape of the negative, L, which will give exact objective tone reproduction with the positive material, P; and the curve shape of the negative, O, which will give proportional tone reproduction with the positive material, P.

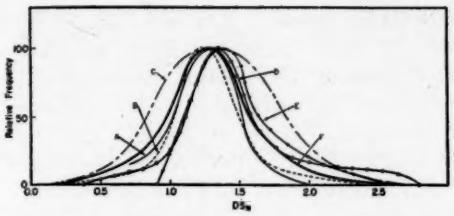


FIG. 14. Frequency of occurrence of different values of density scale for various groups of negatives. Curves *A*, *B*, *C*, and *D* apply to groups of amateur-type negatives. Curve *E* applies to a group of studio portrait negatives. Curve *F* represents the average for all of the amateur-type negatives.

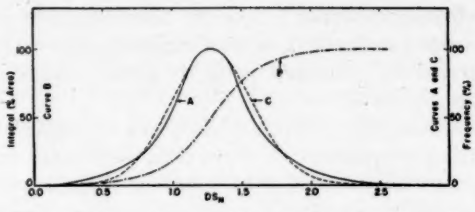


FIG. 15. Curve *A* is a repetition of curve *F* of Fig. 14 and curve *B* is its first derivative. Curve *C* is a Gaussian frequency distribution curve for comparison with curve *A*.

agree upon a method for specifying curve shape, that the use of the three critical points, as discussed in the previous paragraphs, be adopted.

V. THE FREQUENCY DISTRIBUTION OF THE DENSITY SCALES OF NEGATIVES

Experimental data are available regarding the density scales of negatives printed by amateurs, photo-finishers, and professional photographers. The curves in Fig. 14 show the frequency of occurrence as a function of the density scales, DS_n. Curve *A* is derived from 500 amateur negatives obtained and examined by C. M. Tuttle[8] in connection with his work on automatic printing. Curve *B* is for 170 negatives made by H. R. Condit and used in the work by Jones and Condit[4] on the luminance characteristics of exterior scenes and in the work of Jones and Nelson[1] on the control of photographic printing. Curve *B*, in fact, represents the negatives the density scales of which were plotted in Fig. 1 of the present paper. Curve *C* of Fig. 14 applies to 292 amateur negatives measured at two amateur photo-finishing stations in Rochester, New York. Curve *D* applies to 100 amateur negatives measured at a different time at another photo-finishing station in Rochester, New York. Curve *E* is for 150 portrait negatives obtained from a portrait studio. With the exception of the last group, all these negatives may be considered as typical of the amateur field.

The heavy curve, *F*, of Fig. 14 is the weighted average obtained by using all the data except that applying to the portrait negatives. This curve then is based on 1062 negatives which are likely to be fairly representative of amateur practice in this country.

These negatives were all measured on a densitometer which was calibrated in terms of diffuse transmission density. It was capable of reading small areas on the negatives, the diameter of the reading aperture being 0.2 mm. The maximum and minimum densities were measured on each negative, with careful exploration of the negative to make certain that the true maximum and minimum values were obtained. The density scale of each negative was obtained from the relation:

$$DS_n = D_{\max} - D_{\min}. \tag{15}$$

It is interesting to observe how closely the average frequency distribution curve, *F*, of Fig. 14 approaches a normal Gaussian function. For convenient comparison, the average curve has been replotted as curve *A* in Fig. 15. The integral of curve *A* is shown as curve *B*. The dotted curve, *C*, of Fig. 15 is a Gaussian distribution which has been adjusted to what seems to be the best fit to our experimental data. This curve is represented by the equation,

$$y = y_o \exp[-(D_m - D_i)^2/2\sigma], \tag{16}$$

in which σ has a value of 0.297.

One test for normality of a frequency distribution curve has been described by Selwyn.[9] The test consists in plotting its integral curve on arithmetic probability paper and noting how closely it fits a straight line. This has been done for curve *A* of Fig. 15 and the result is shown as curve DS_n in Fig. 16. The straight line labeled "normal" in Fig. 16 corresponds to the Gaussian frequency distribution, curve *C*, of Fig. 15. It is

[8] C. M. Tuttle, "Photoelectric photometry in the printing of amateur negatives," J. Franklin Inst. **224**, 315 (1937).

[9] E. W. H. Selwyn, "Note on the use of probability paper," Proc. Phys. Soc. **50**, 914 (1938).

obvious that the frequency distribution of the density scales of these 1062 negatives approaches but does not agree perfectly with a Gaussian distribution.

Although a closer approach to a Gaussian distribution might occur if the data were based on a much larger number of amateur negatives obtained from various parts of the country, it seems likely that some difference would nevertheless exist. In the case of the portrait negatives, curve *E* of Fig. 14, there were no low values of density scale and the departure from the Gaussian distribution is considerable. These negatives were made in a professional portrait studio. In this type of work the lighting and exposure are more carefully adjusted and controlled than they are in amateur photography.

The factors which determine the density scales of the negatives are the luminance scales of the scenes, the gradient characteristics of the negative material, the amount of camera flare arising from the luminance distribution in the scene and its relation to the flare characteristics of the cameras, the negative exposure, and the negative development. In amateur photography, an appreciable number of negatives may be underexposed or overexposed, but the majority of negatives are not. The negative development is usually such that the gamma of the negative material lies between 0.75 and 0.95. It appears probable from these considerations that the major cause of the variation in density scales of negatives is the variation in the luminance scales of the scenes. Further evidence on this point is obtained by referring to direct measurements by these Laboratories of the luminance scales of exterior scenes as reported by Jones and Condit.[4] The luminance scales of these scenes varied from 27 to 760, the average being 160. The negatives made of these scenes, at the time of the measurements, were given normal exposure and normal development as described earlier in this paper but, nevertheless, the density *scales* of these negatives *vary* widely. The frequency of occurrence of the density scales of these negatives is shown as curve *B* (dotted) of Fig. 14. This curve is not as broad as curve *F*, which includes the negatives measured at the photo-finishing stations, and this is evidence that exposure and processing variations do contribute materially to the variations in negative density scale. The point we wish to emphasize, however, is that, even though exposure and processing were controlled perfectly, there would, nevertheless, exist a large variation in the density scales of the negatives in the amateur field because of the variations in the luminance scales of scenes.

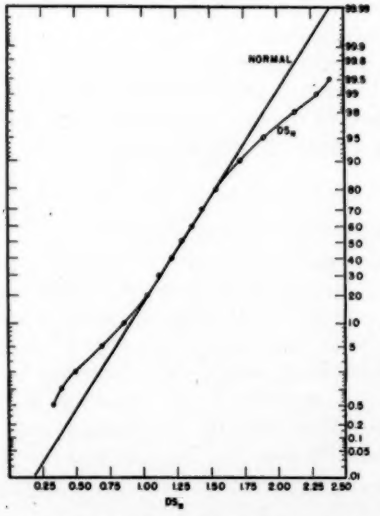


FIG. 16. Test for normality of the frequency distribution curve, *F*, of Fig. 14.

The frequency of occurrence of the density scales of negatives is of particular significance in explaining the *relative amounts* of the different grades of paper consumed by photographers and photo-finishers. It is to be expected from a consideration of the shape of the frequency distribution curve, *F*, in Fig. 14 and from the relation between negative density scale and paper log exposure scale discussed in Sections I and II of this paper that a large amount of what may loosely be termed the "medium" grades of paper should be required and only a small amount of the hard and soft grades. An examination of the sales volume of the different grades of paper shows that this does occur.

VI. SUMMARY AND CONCLUSIONS

1. In photographic printing, the density scale of the negative can be used as a means of choosing the grade of paper for making the print. The proposed method is derived from further analysis of the correlation which, in our previous

communication,[1] was shown to exist between the *density scales* of negatives and the *exposure scales* of papers. A number of negatives of different scenes were printed, each on several grades of paper, and the best grade for each negative was determined by qualified judges who selected the *first-choice* prints. It was found that, because of the influence of the brightness distribution and subject matter in the scenes photographed, an accurate prediction cannot always be made of the exposure scale of the paper which will give a *first-choice* print from a negative of known density scale. However, a careful rejudging of the prints resulted in a formula, intended as a guide in printing, which gives an excellent yield of high quality prints even though it does not give a *first-choice* print in all cases. This formula is as follows:

$$TP = 1.35\left[\frac{\log SES(a) + \log SES(a+1)}{2}\right] - 0.40, \quad (17)$$

in which $SES(a)$ and $SES(a+1)$ are the *sensitometric exposure scales* of any two adjacent grades of paper and TP is the *transition point* or *negative density scale* value at which the transition from one grade of paper to the next should be made. For example, if a paper is available in three grades (soft, medium, and hard), two transition points are calculated. Negatives having density scales lying between these two transition points would then be printed on the medium paper; negatives having density scales greater than the greater transition point would be printed on the soft paper; and negatives having density scales less than the lesser transition point would be printed on the hard paper.

2. The *sensitometric exposure scale* (*SES*) values used in the above equation are determined from the $D-\log E$ curves of the papers by means of the fractional-gradient criterion ($0.1\bar{G}$) defined in our previous paper and adopted by the American Standards Association.[2] Thus, *SES* is the American Standard exposure scale. Values of *SES* or log*SES* are usually published by or are available from the paper manufacturer. For accurate work it is necessary to determine the *SES* values of the particular batches of paper as they are being used with the worker's particular printing equipment and processing conditions.

3. The *density scales* of the negatives may be obtained by measuring the maximum and minimum densities of the negatives with a densitometer which is capable of reading small areas. If the densitometer reads diffuse density, the equation given is directly applicable to contact printing. The equation is also directly applicable to projection printing when the effective densities of the negative are measured by an accurate easel photometer. It is also applicable in many cases to projection printing even when densities are measured by a diffuse densitometer, since many projection printers are only slightly specular. The flare light in such printers is often sufficient to offset the specularity to the extent that the effective printing densities of the negative are very nearly equal to diffuse densities. For projection printers which depart appreciably from this condition, a correction of the density values must be made, but it is outside the scope of this paper to deal with this aspect of the problem.

4. The term *useful exposure scale* has been abandoned. Originally the term was intended to apply to a numerical value which would be assigned to each paper to indicate the average of the density scales of the negatives which give *first-choice* prints on the paper. A review of the data shows that this average is not a unique function of each paper alone but depends upon the spacing, with respect to log*SES*, of the consecutive grades of paper in the group and also upon the frequency distribution of the density scales of the negatives. Thus, no fixed numerical value of this characteristic can be assigned to each paper. Moreover, the term is unnecessary, since, as shown in items 1, 2, and 3, the problem can be solved without reference to this characteristic.

5. The rejudgment of the prints, in order to obtain a unique relation between the density scale of the negative (DS_n) and the sensitometric exposure scale of the paper (*SES*), revealed that for optimum results the following general principles apply:

a. For the soft papers, DS_n should, in most cases, exceed log*SES*.
b. For the medium papers, DS_n should, on the average, be equal to log*SES*.
c. For the hard papers, DS_n should, in most cases, be less than log*SES*.

These principles are incorporated in the equation given in item 1. Sanders[3] has pointed out that these principles can also be derived (but only in an approximate or a qualitative form) from an analysis of our correlation diagram (Fig. 1), in which the statistical relation between $\log SES$ and DS_n for *first-choice* prints is shown. His reasoning is based on established methods of statistical mathematics used in analyzing correlation data and has contributed to a better understanding of the statistical nature of the problem.

6. Sanders[3] has suggested that, because of the principles described in item 5, it may be of value to define a new kind of exposure scale which would be greater than *SES* for the soft papers, equal to *SES* for the medium papers, and less than *SES* for the hard papers, by the amounts indicated by our statistical data. We suggest that the term *transition point index*, I_{tp}, be applied to this aspect of exposure scale because it would be used in computing the transition point for any two adjacent grades of paper simply by averaging the I_{tp} values for the two papers. Since I_{tp} cannot be obtained directly from the $D-\log E$ curve of the paper because the highlight end of the interval sometimes extends beyond the point of zero gradient on the toe of the curve, it is necessary to derive it from the sensitometric exposure scale by means of the equation:

$$I_{tp}=1.35\ \log SES_y-0.40. \qquad (18)$$

This equation and the one in item 1 give identical results in printing. They are, in fact, equivalent relations; only the order of the operations in computing the desired transition points have been changed. The choice between them depends upon the nature of the practical application.

7. The proposed method for choosing the grade of paper from the sensitometric exposure scale (or from the transition point index) of the paper gives prints which have somewhat lower contrast than the prints ordinarily supplied by the average photo-finisher in this country. In the opinion of our judges, these lower contrast prints are of better quality.

8. We suggest that the *sensitometric exposure scale* be adopted as the basis for determining the grade of photographic paper. The term *scale index* is suggested for designating the grades so that they will not be confused with the older grade numbers. The formula,

$$\text{scale index}=100\ \log SES,$$

is proposed for determining these values, where *SES* is the sensitometric exposure scale defined in our previous paper[1] and adopted by the American Standards Association[2] as an American Standard exposure index. The *contrast* of the paper is not satisfactory as a basis for deriving grade numbers because, for papers which differ in $D-\log E$ curve shape or in maximum density, it gives rise to equal grade numbers for papers which are unsuitable for printing the same negatives or unequal grade numbers for papers which are suitable for printing the same negatives. While the term *contrast grade number* is sometimes used, it is not advisable to continue this usage since the grade number of a paper is related to the exposure scale and not to the contrast of a paper, the exposure scale being only one component of contrast.

9. For specifying the shapes of the $D-\log E$ curves of photographic papers, Morrison's[7] method is recommended.

10. The frequency distribution curve for the density scales of amateur negatives has been shown to be similar to a Gaussian or normal frequency distribution curve but the two are not identical in shape. The majority of negatives have density scales such that they require the medium grades of paper.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 11 NOVEMBER, 1948

On the Theory of Imperfect Diffraction Gratings

U. FANO
X-Ray Section, National Bureau of Standards, Washington, D. C.
(Received July 28, 1948)

The "dynamical" theory of gratings originally developed by Rayleigh and Voigt is applied to derive the intensity of the light diffracted in various directions by an imperfect grating of finite area. The problem is reduced to the numerical solution of a system of linear equations by an approximation method in which "ghosts" and high order spectra are treated as perturbations of the main spectra. Current elementary theories are then seen to yield merely order of magnitude estimates of the intensity of the ghosts caused by various grating imperfections.

I. INTRODUCTION

THE theory of the light distribution in the spectra produced by gratings has followed primarily Fresnel's and Kirchhoff's elementary approach to the theory of diffraction through the application of Huygens' principle. A simple theory of this type assumes that the light diffracted in different directions by a series of grooves is the same as that emitted in those directions by a series of infinitely thin linear coherent sources of light placed in the center of each groove. Such a theory is capable of predicting the effect of a great variety of imperfections of a grating[1] by representing each type of imperfection as an anomaly of the amplitude or phase of the various light sources. The same results are obtained by representing the diffracted light reaching any point as the superposition of waves traveling from the real source in a straight line to each point of the grating surface and thence, without any change in amplitude or phase, again in a straight line to the point of observation.[2]

The most important results of theories of this type are no doubt correct, but it is realized that they are not consistent or complete, as they do not consider the actual boundary conditions at the surface of the grating. As a result, the intensity of the light diffracted in different directions is found to be independent of the nature of the grating's surface and of the polarization of the impinging light, contrary to expectation and experience. Many applications of the Fresnel-Kirchhoff theory of diffraction have been rederived as approximations to a consistent theory.[3] This does not seem to have been done for the theory of diffraction gratings. Hence there is no clear guide as to which detailed results of the elementary theory can be trusted to be correct and how far such a theory should be developed in its application to particular cases.

Rayleigh,[4] and then Voigt,[5] set forth a consistent "dynamical theory of gratings," but this theory has thus far been applied only to perfect, infinitely extended, gratings. Furthermore, its practical application seemed to require the rather unrealistic assumption that the depth of the grating's grooves be small as compared to the wave-length of the light. Earlier investigations[6,7] have shown that the Rayleigh-Voigt method is an effective tool for applications to "anomalous" gratings. An attempt was therefore made to apply it to the case of imperfect gratings.[8]

This attempt has progressed to the point where the way seems open to carry out a consistent numerical application of the theory. It has, in fact, turned out that the very existence of imperfections of the grating makes it possible to introduce a new more realistic approximation method for determining the amplitude of the diffracted waves. No effort has been made, however, to develop this approximation method

[1] E.g., C. M. Sparrow, Astrophys. J. **49**, 65 (1919).
[2] E.g., H. A. Rowland, *Physical Papers* (Johns Hopkins University Press, Baltimore, 1902), p. 525.
[3] E.g., P. Frank and R. V. Mises, *Differentialgleichungen der Physik* (Friedrich Vieweg and Sohn, Braunschweig, 1935), Vol. 2, p. 853 ff.
[4] Lord Rayleigh, Proc. Roy. Soc. (A) **79**, 399 (1907).
[5] W. Voigt, Göttinger Nachrichten 40 (1911).
[6] U. Fano, Ann. d. Physik **32**, 393 (1938).
[7] U. Fano, J. Opt. Soc. Am. **31**, 213 (1941).
[8] A detailed analytical treatment by this method has also been given by C. T. Tai, for the case in which the grating spacing is much shorter than the radiation wave-length, see Tech. Rep. No. 28, Cruft Laboratory, Harvard University, Cambridge, Massachusetts, Jan. 15, 1948.

analytically. Similarly, the results described in this paper throw some new light on the significance of the elementary theories, but perhaps not yet as much as would be desirable.

II. QUALITATIVE TREATMENT

The Rayleigh-Voigt theory deals with a plane grating which is assumed to consist of an infinitely extended ruled metal surface. It aims at calculating the light diffracted by such a grating when the incident light consists of a plane wave traveling in a direction perpendicular to the grooves. The diffracted light consists of the discrete set of plane reflected and transmitted waves which is predicted by elementary theories, and constitutes the spectra of different orders.[9] These waves appear to an observer to be due to "virtual" light sources located at infinite distance on the opposite side of the grating surface with respect to the observer. The amplitudes and phases of these sources are determined in such a way that the whole radiation field fulfils the boundary conditions at the grating surface. These conditions can be put in the form of a system of infinitely many linear algebraic equations:

$$\sum_{n=-\infty}^{\infty} c_{n'n}x_n = d_{n'},^{10} \qquad (1)$$

where the c's are Fourier coefficients of certain functions of the grating profile, the d's are similar coefficients proportional to the amplitude of the incident wave, and the x's are the unknown amplitudes of the diffracted waves. It is the periodicity of the grating surface which causes the Fourier coefficients, and hence the diffracted waves, to form discrete sets. This property of periodicity is absent, strictly speaking, when the grating is imperfect and when the ruling extends over a limited area. Integral equations and continuous sets of diffracted waves are then met in this realistic case. It will now be shown how the resulting difficulties can be overcome.

The light diffracted by a limited ruled grating area is, in the main, independent of the nature and shape of the materials—such as the frame—which surround that area. Only the light diffracted by the surface portions very near the edge of the grating will depend on the characteristics of both the grating and the surrounding materials; this amount of light is quantitatively negligible. It is therefore a good approximation to assume that the grating is framed into an opening within a black mask.

As was pointed out before, the presence of a perfect diffraction grating surface of infinite extension is equivalent for an observer—whether he be located on the air or the metal side of the grating surface—to the presence of a set of virtual light sources located on the opposite side of the grating surface with respect to him. An observer viewing a limited grating surface will then have very approximately the same impression as though he were viewing an infinite grating which is covered by a black mask except for a limited area, i.e., as though he were viewing all the virtual light sources pertaining to an infinite grating through the opening in the mask. (This imaginary infinite grating will be called the "fictitious grating".) The light reaching the observer from each virtual source will then no longer be parallel but spread out into a Fraunhofer diffraction pattern as light which has passed through the opening of a mask. Hence, while monochromatic light diffracted from the fictitious grating would be displayed on a spectroscope screen as a series of infinitely sharp "lines," the light diffracted by the actual grating will show in the place of each line a broader pattern, consisting of a principal fringe of finite width flanked by two infinite sets of "satellites" of decreasing intensity. This result is known from the elementary theory, but its significance may not have been adequately stressed.

Any assumption whatsoever can be made concerning that portion of the fictitious infinite grating which is covered by the mask. Advantage can be taken of this arbitrariness for the treatment of an imperfect grating. It can thus be assumed that the fictitious grating consists of an infinite set of identical patches, all equal to the actual grating area surrounded by the mask. Since the patches are identical, the fictitious grating is perfectly periodic and the ordinary Rayleigh-Voigt theory can be applied to it. The

[9] This set includes waves reflected and transmitted in "complex directions," i.e., waves confined to the proximity of the grating's surface. See reference 7, p. 214. For a derivation of the composition of the diffracted light, see reference 6, p. 402.

[10] See reference 7, Appendix.

periodicity of this fictitious grating is, however, not the microscopic periodicity of the individual grooves (this would rather be a "near periodicity" in view of imperfections), but a macroscopic periodicity of ruled patches. Without even taking into account the fact that each patch is itself a grating, even though an imperfect one, it follows that the diffraction spectrum of the fictitious grating without a mask consists of very closely spaced infinitely sharp "lines," that is, of waves progressing in a series of closely spaced but sharply defined directions. (The angle of emergence ϑ_q of the diffracted wave of qth order is given by the usual formula $\sin\vartheta_q - \sin\vartheta_0 = q\lambda/L$, where λ is the wave-length and L the width of the patch. Since λ/L is so very small, the difference of ϑ_{q+1} and ϑ_q is very small too.)

Consideration should now be given to the fact that each patch of the fictitious grating consists of a number, N, of nearly identical grooves. From here on two different periodicities will have to be considered, namely, the "patch periodicity" and the "groove periodicity" within the patches. A spectrum of qth order of the patch periodicity may be considered as a spectrum of nth order of the groove periodicity, with $n = q/N$, since there are N grooves to a patch. If, now, the groove ruling were perfect so that the system of ruled patches in the fictitious grating would blend into a single perfectly periodic infinite set of grooves, the periodicity of the patches as such would lose any significance. As a result only the spectra of Nth, $2N$th, $3N$th . . . order of the fictitious grating (i.e., of the patch periodicity) would appear, which coincide with the spectra of 1st, 2d, 3d . . . order of the groove periodicity. (This can also be seen by considering that the whole fictitious grating cannot possibly diffract light in directions towards which each individual patch would diffract none.) Small *systematic* departures from groove periodicity within each patch will, on the contrary, cause the appearance of low intensity spectra of certain orders of the patch periodicity other than a multiple of N, i.e., of fractional orders of the groove periodicity. These spectra are the so-called "ghosts." Thus, for example, a small systematic equal displacement of every third groove from its regular position will cause the appearance of spectra of orders which are multiples of $N/3$ of the patch periodicity. Small *accidentally* distributed departures from periodicity will, on the contrary, cause very faint spectra to appear in all orders and a corresponding weakening of the main spectra. This weakening may affect the low order spectra of the groove periodicity to a negligible extent, but it will completely suppress the spectra of sufficiently high order of that periodicity. This is well known from the elementary theory[11] and is due to the fact that high order spectra stem from high overtones of the groove periodicity and these are in turn suppressed even by small random errors, e.g., in the position of the grooves. The elimination of the high order spectra reduces the system of Eqs. (1) from infinite to finite size and hence makes it accessible to numerical solution.

There remains to be carried out the passage from the spectrum of the fictitious grating to that of the actual, single-patch grating. This is accomplished, as previously shown, by assuming that all but one patch of the fictitious grating is covered by a black mask. Hence each sharp line of the spectrum of the fictitious grating must be replaced by the Fraunhofer diffraction pattern corresponding to the opening in the mask. The angular width of the principal fringe of this diffraction pattern is equal to the ratio of the wave-length λ to the width L of the opening in the mask. Hence the points of zero light intensity flanking the main fringe of the spectrum of qth order of the patch periodicity coincide with the centers of the patterns of the adjacent spectra of $(q-1)$th and $(q+1)$th order. The superposition of these Fraunhofer diffraction patterns of adjacent orders will therefore yield a complicated pattern, *unless* the spectrum of a certain order is much more intense than all the adjacent ones. The latter case is that which occurs in practice for nearly perfect gratings.

In the first place, as previously stated, the spectra of Nth, $2N$th, $3N$th . . . order of the patch periodicity are by far the most intense ones, since they alone would be present if the groove periodicity were perfect. Hence the adjacent spectra of $(N-2)$th, $(N-1)$th, $(N+1)$th, $(N+2)$th . . . , etc., orders will be far less intense than the main spectra and only cause a distortion of their line shape from the perfect

[11] See reference 1, p. 83. See also reference 14.

Fraunhofer pattern. These spectra of orders adjacent to multiples of N are those whose appearance is caused by the so-called "errors of run" in the ruling.

In the second place small periodic errors of the ruling, e.g., errors affecting every mth groove, cause the appearance of weak spectra of orders N/m, $2N/m$, etc., of the patch periodicity. These spectra are still far more intense than the adjacent spectra of orders $N/m-1$, $N/m+1$, etc., and hence will give rise to rather faint nearly perfect Fraunhofer patterns, the so-called "ghost lines."

Finally, the Fraunhofer patterns of the faint spectra of all orders whose appearance is caused by accidental errors, when superimposed on one another, will not give rise to any characteristic light distribution but to a continuous "gray background."

The qualitative analysis carried out thus far indicates that the amplitude of the light diffracted in different directions by a finite imperfect grating should be conveniently represented as a superposition of Fraunhofer diffraction patterns corresponding to an opening in a mask shaped like the ruled area of the grating. The amplitude of each pattern is characterized by a constant factor which is interpreted as the amplitude of one of the sharp lines diffracted by an infinite fictitious grating and hence must be given by the solution of the Rayleigh-Voigt system of Eqs. (1) for that grating. Finally, this system should be solved by an approximation method in which the spectra of high integral order and of all fractional orders of the groove periodicity are considered as perturbations of the main spectra.

III. CALCULATIONS

The explicit formulation of the system of Eqs. (1) has been given previously, and hence the application of the method indicated above should offer no difficulty. The analytical development will, however, be reported here for convenience.

The qualitative arguments of the previous section will be utilized merely as a guide to choose the form of the solution, and the formulation of the problem will not require any undue schematization. In this formulation it seems more convenient to specify that the ruled grating area is framed in a reflecting rather than in a black mask. For the sake of definiteness it will be assumed that half the space is occupied by a material which will be referred to as "the metal," and the other half by another material which will be referred to as "air." The surface separating these materials is assumed to be a plane, namely, the xy coordinate plane, except for a ruled portion, namely, the grating, which extends from $x=0$ to $x=L$. The ruling will be assumed to be perfectly parallel to the y axis and to extend over an infinite length in this direction. This assumption simplifies the calculation since the problem is thereby reduced to a two-dimensional one in the xz plane. It should not introduce any essential limitation since the extension of the following procedure to a more realistic case seems straightforward. Since the ruling extends only from $x=0$ to $x=L$, the surface of the grating is described by an equation $z=\zeta(x)$ in which ζ vanishes when x is outside the limits 0 and L.

If the electric or the magnetic field of the incident wave is parallel to the y axis, the same will be true for the diffracted waves; the radiation amplitude can then be treated simply as a scalar quantity. The boundary conditions at the surface of the grating will, however, be different, depending on which field is parallel to the y axis. A general type of polarization can always be resolved into a superposition of these two simple types of polarization.

The amplitude of the radiation will be represented as:

$$\begin{aligned} &A(x, z)\exp(-2\pi i\nu t) \text{ in air } (z \geqslant \zeta(x)),\\ &B(x, z)\exp(-2\pi i\nu t) \text{ in the metal } (z \leqslant \zeta(x)). \end{aligned} \tag{2}$$

The functions A and B must fulfil the boundary conditions at the surface of the grating, namely:

$$A(x, \zeta)=B(x, \zeta),$$

$$K\left(\frac{\partial A}{\partial z}-\frac{\partial A}{\partial x}\frac{d\zeta}{dx}\right)_{z=\zeta}=\left(\frac{\partial B}{\partial z}-\frac{\partial B}{\partial x}\frac{d\zeta}{dx}\right)_{z=\zeta}, \tag{3}$$

where $K=1$ if the electric field is parallel to the y axis, $K=n^2$, i.e., the squared complex refractive index of the metal, if the magnetic field is parallel to it. The two polarizations can thus be treated simultaneously.

The radiation field will consist of an incident wave and of diffracted outgoing waves proceeding in various directions from the surface of

the grating. It is then convenient to represent the radiation fields A and B in (2) as superpositions of plane waves going in different directions; the determination of their amplitudes is the object of the calculation. A plane wave propagating away from the grating surface is represented in air and in the metal, respectively, by

$$\exp[2\pi i(ux+vz)] \quad \text{and} \quad \exp[2\pi i(ux-wz)], \tag{4}$$

where u and v are the components of the wave vector in air, u and w the components in the metal. Calling k the length of the wave vector in air, i.e., the wave number, we must have:

$$v=v(u)=(k^2-u^2)^{\frac{1}{2}}, \quad w=w(u)=(n^2k^2-u^2)^{\frac{1}{2}}. \tag{4'}$$

In order that the amplitude of these waves be finite and that their direction be outwards, the arguments of the complex quantities v and w must fall between 0 and $\pi/2$:

$$0\leqslant \arg(v)\leqslant \pi/2, \quad 0\leqslant \arg(w)\leqslant \pi/2. \tag{4''}$$

The whole radiation field is then represented in the form

$$A(x,z)=I\exp[2\pi i(u_0x-v_0z)]+\int_{-\infty}^{\infty} duR(u)\exp[2\pi i(ux+vz)], \tag{5}$$

$$B(x,z)=\int_{-\infty}^{\infty} duT(u)\exp[2\pi i(ux-wz)],$$

where $R(u)$ and $T(u)$ are the amplitudes of the waves reflected and transmitted in the directions corresponding to all different values of u. The negative sign in front of v_0[12] indicates that the incident wave of amplitude I is traveling towards the surface of the metal. $R(u)$ and $I(u)$ must be determined by solving Eqs. (3), which can now be written in the form

$$\begin{aligned}&\int_{-\infty}^{\infty} du\exp(2\pi iux)\\&\quad\times[R(u)\exp(2\pi iv\zeta)-T(u)\exp(-2\pi iw\zeta)]\\&\qquad=-I\exp[2\pi i(u_0x-v_0\zeta)].\end{aligned} \tag{3'}$$

$$\begin{aligned}&\int_{-\infty}^{\infty} du\exp(2\pi iux)\\&\quad\times[K(v-u\,d\zeta/dx)R(u)\exp(2\pi iv\zeta)\\&\quad+(w+u\,d\zeta/dx)T(u)\exp(-2\pi iw\zeta)]\\&\qquad=K(v_0+u_0\,d\zeta/dx)I\exp[2\pi i(u_0x-v_0\zeta)].\end{aligned}$$

Since the function $\zeta(x)$ vanishes outside the interval from 0 to L, the exponential functions of ζ can be represented by means of a Fourier series and of the "pulse" function $P(x)$ which is equal to 1 for $0\leqslant x\leqslant L$ and vanishes otherwise. Thus, for example,

$$\begin{aligned}&\exp[2\pi iv\zeta(x)]=1-P(x)+P(x)\sum_{p=-\infty}^{\infty} a_p(v)\exp(2\pi ipx/L),\\&(d\zeta/dx)\exp[2\pi iv\zeta(x)]\\&\qquad=P(x)\sum_{p=-\infty}^{\infty} a_p(v)(p/vL)\exp(2\pi ipx/L),\end{aligned} \tag{6}$$

where

$$a_p(v)=(1/L)\int_0^L \exp[2\pi i(v\zeta(x)-px/L)]dx.$$

These and similar expression can be introduced in Eqs. (3'), after which these equations should be multiplied by $\exp(-2\pi iu'x)$ and integrated over x from $-\infty$ to ∞. This means replacing Eq. (3') valid for all values of x by equations among the coefficients of their Fourier integral expansions.

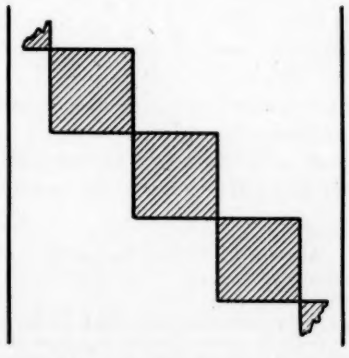


FIG. 1. Disposition of the elements of the matrix $c_{n'n}$. The large elements are confined within the shadowed areas.

[12] An index applied to v or w means that these functions have the values corresponding to the value of u with the same index.

Using the formulae

$$\int_{-\infty}^{\infty} \exp[2\pi i(u-u')x]dx = \delta(u-u'),$$

$$\int_{-\infty}^{\infty} f(u)\delta(u'-u)du = f(u'),$$

$$\int_{-\infty}^{\infty} \exp[2\pi i(u-u')x]P(x)dx = \int_0^L \exp[2\pi i(u-u')x]dx$$

$$= \exp[i\pi(u-u')L]\frac{\sin[\pi(u-u')L]}{\pi(u-u')} = \Delta(u-u'),$$

(where the last function represents the amplitude distribution in the Fraunhofer diffraction pattern of a slit of width L), Eq. (3′) yields

$$\begin{aligned}
&[R(u')-T(u')] \\
&\quad -\int_{-\infty}^{\infty}[R(u)-T(u)]\Delta(u'-u)du \\
&\quad +\sum_{p=-\infty}^{\infty}\int_{-\infty}^{\infty}[a_p(v)R(u)-a_p(-w)T(u)] \\
&\qquad \times\Delta(u'-u-p/L)du \\
&\quad = -I[\delta(u'-u_0)-\Delta(u'-u_0) \\
&\qquad +\textstyle\sum_p a_p(-v_0)\Delta(u'-u_0-p/L)]; \\
&[Kv'R(u')+w'T(u')] \\
&\quad -\int_{-\infty}^{\infty}[KvR(u)+wT(u)]\Delta(u'-u)du \\
&\quad +\sum_{p=-\infty}^{\infty}\int_{-\infty}^{\infty}[K(v-up/vL)a_p(v)R(u) \\
&\quad +(w-up/wL)a_p(-w)T(u)] \\
&\qquad \times\Delta(u'-u-p/L)du \\
&\quad = I\{Kv_0[\delta(u'-u_0)-\Delta(u'-u_0)] \\
&\quad +\textstyle\sum_p K(v_0-u_0p/v_0L) \\
&\qquad \times a_p(-v_0)\Delta(u'-u_0-p/L)\}.
\end{aligned} \tag{7}$$

The various groups of terms of these equations correspond to the different terms of the first expansion (6).

The equations thus obtained represent a general formulation of the problem. At this point the considerations of the preceding section are utilized to assign a specific form to the unknown amplitudes $R(u)$ and $T(u)$, namely,

$$\begin{aligned}
R(u) &= \rho[\delta(u-u_0)-\Delta(u-u_0)] \\
&\quad +\sum_{q=-\infty}^{\infty} r_q\Delta(u-u_0-q/L), \\
T(u) &= \tau[\delta(u-u_0)-\Delta(u-u_0)] \\
&\quad +\sum_{q=-\infty}^{\infty} t_q\Delta(u-u_0-q/L).
\end{aligned} \tag{8}$$

The last group of terms in each expression represents the amplitudes of the waves diffracted by the grating, described as a set of superposed Fraunhofer diffraction patterns. As explained at the end of Section II, the coefficients r_q and t_q represent the amplitudes of the waves that would be diffracted by the infinite "fictitious" grating. The presence of the first two terms in each expression is due to the fact that we have described the grating as surrounded by a plane reflecting surface rather than by a black mask. The first term represents the light regularly reflected or refracted by the plane surface, the second term represents the diffraction effect resulting from the exclusion of the ruled position from the surface.

The expression (8) can now be introduced in Eqs. (7). Here an approximation will be used, by assuming that a slowly varying function of u multiplied by the function $\Delta(u-u'')$ which is very small unless $u\sim u''$ may be replaced by the value of the same function at $u=u''$. This assumption is equivalent to that made in Section II that the light diffracted by the grating is not affected by the materials surrounding the grating except for a negligible edge effect.[13]

The assumption just introduced makes it possible to remove the coefficients a_p, $(v-up/vL)$, and $(w-up/wL)$ from under the signs of integration in (7). Use will be made of the identity

$$\int_{-\infty}^{\infty}\Delta(u-u')\Delta(u''-u)du = \Delta(u''-u'), \tag{9}$$

[13] In fact, if no assumption were made at this point, the calculation could be carried further, but more laboriously, taking into account interaction terms between the light diffracted from the grating and that coming from the rest of the metal surface. In that case r_q and t_q should still be considered as slowly varying functions of u.

from which it follows that

$$\int_{-\infty}^{\infty} R(u)\Delta(u''-u)du = \sum_q r_q\Delta(u''-u_0-q/L), \quad (9')$$

together with the corresponding formula for $T(u)$. The symbol $u_0^{(q)}$ will be used to indicate u_0+q/L, and, similarly, the symbol $v_0^{(q)}$.[11] Substitution of (8) into (7) yields then:

$$\begin{aligned}
&(\rho-\tau)[\delta(u'-u_0)-\Delta(u'-u_0)]\\
&\quad+\sum_q (r_q-t_q)\Delta(u'-u_0^{(q)})\\
&\quad-\sum_q (r_q-t_q)\Delta(u'-u_0^{(q)})\\
&\quad+\sum_p\sum_q [a_p(v_0^{(q)})r_q-a_p(-w_0^{(q)})t_q]\\
&\qquad\times\Delta(u'-u_0^{(p+q)})\\
&=-I[\delta(u'-u_0)-\Delta(u'-u_0)\\
&\qquad+\sum_p a_p(-v_0)\Delta(u'-u_0^{(p)})];\\
&(Kv_0\rho+w_0\tau)[\delta(u'-u_0)-\Delta(u'-u_0)] \qquad (10)\\
&\quad+\sum_q (Kv_0^{(q)}r_q+w_0^{(q)}t_q)\Delta(u'-u_0^{(q)})\\
&\quad-\sum_q (Kv_0^{(q)}r_q+w_0^{(q)}t_q)\Delta(u'-u_0^{(q)})\\
&\quad+\sum_p\sum_q [K(v_0^{(q)}-u_0^{(q)}p/v_0^{(q)}L)a_p(v_0^{(q)})r_q\\
&\quad+(w_0^{(q)}-u_0^{(q)}p/w_0^{(q)}L)a_p(-w_0^{(q)})t_q]\\
&\qquad\times\Delta(u'-u_0^{(p+q)})\\
&=IK\{v_0[\delta(u'-u_0)-\Delta(u'-u_0)]\\
&\quad+\sum_p (v_0-u_0p/v_0L)a_p(-v_0)\Delta(u'-u_0^{(p)})\}.
\end{aligned}$$

If then the coefficients of $\delta(u'-u_0)-\Delta(u'-u_0)$ and of the various functions $\Delta(u'-u_0^{(p)})$ are separately equated to each other, the final systems of algebraic equations are obtained.

$$\rho-\tau=-I; \quad Kv_0\rho+w_0\tau=I; \qquad (11a)$$

$$\begin{aligned}
&\sum_q [a_{p-q}(v_0^{(q)})r_q-a_{p-q}(-w_0^{(q)})t_q]\\
&\qquad=-Ia_p(-v_0);\\
&\sum_q [K(v_0^{(q)}-(p-q)u_0^{(q)}/v_0^{(q)}L)a_{p-q}(v_0^{(q)})r_q \qquad (11b)\\
&+(w_0^{(q)}-(p-q)u_0^{(q)}/w_0^{(q)}L)a_{p-q}(-w_0^{(q)})t_q]\\
&\qquad=IK(v_0-pu_0/v_0L)a_p(-w_0).
\end{aligned}$$

The system (11a) is the usual one which leads to the Fresnel formulae for the amplitude of the light regularly reflected and refracted on the plane portion of the metal surface. It is of no further interest once it is shown that the amplitudes of these waves separate out from the amplitudes of the waves diffracted by the grating.

The system (11b) constitutes the formulation of the diffraction problem for the "fictitious" grating and can be represented schematically in the form (1) by taking, for example,

$$\begin{aligned}
&x_{2q}=r_q; \qquad x_{2q+1}=t_q;\\
&c_{2p,\,2q}=a_{p-q}(v_0^{(q)}); \quad c_{2p,\,2q+1}=-a_{p-q}(-w_0^{(q)});\\
&c_{2p+1,\,2q}=K[v_0^{(q)}-(p-q)u_0^{(q)}/v_0^{(q)}L]\\
&\qquad\qquad\times a_{p-q}(v_0^{(q)}); \qquad (1')\\
&c_{2p+1,\,2q+1}=[w_0^{(q)}-(p-q)u_0^{(q)}/w_0^{(q)}L]\\
&\qquad\qquad\times a_{p-q}(-w_0^{(q)});\\
&d_{2p}=-Ia_{2p}(-v_0);\\
&d_{2p+1}=K(v_0-pu_0/v_0L)I.
\end{aligned}$$

If the grating is *perfectly periodic* within the ruled patch and consists of N identical sections of width L/N, all coefficients a_p, except those for which p is a multiple of N (inclusive of 0), vanish, as can be seen from (6). Hence all the elements of the matrix of coefficients $c_{n'n}$ vanish, except those for which the difference $p-q$ is a multiple of N. This matrix can then be arranged in such a way as to reduce to a set of N square submatrices arranged along its main diagonal (see Fig. 1). The s-th square submatrix contains elements whose values of p and q are equal to multiples of N plus a number $0\leqslant s<N$. The system of equations then breaks down into N separate systems, each of which involves only amplitudes r_q, t_q whose indices q are equal to multiples of N plus a number s. Furthermore, the quantities d_n' on the right side of (1) vanish for all systems except that with $s=0$, because of their definition (1'). Hence only amplitudes whose indices q are exact multiples of N will be different from 0, in agreement with Rayleigh's original theory.

In the general, realistic case of imperfect gratings it is still convenient to arrange the matrix of coefficients $c_{n'n}$ as in the case of perfect gratings. The matrix elements outside the diagonal set of square submatrices will then not be identically zero, but the larger matrix elements

will still be concentrated in these submatrices. At the same time the quantities $d_{n'}$, for which p is not a multiple of N, will not be identically zero, but the larger ones will still be those for which p is a multiple of N. Hence it is appropriate to apply a method of approximation in which the "small" coefficients c and d are initially taken to be zero, while their actual values are taken into account in further approximations. It follows then that the amplitudes r_q, t_q whose index q is a multiple of N plus a number s are of an order of magnitude determined by the value of the coefficients a_p whose index p is also a multiple of N plus the same number s. This can be seen, for example, by writing the formal solution of the system of equations in the usual form, i.e., as a ratio of two determinants.

Now, the formula (6′) which determines the *coefficient* a_p has the same structure as the formula which gives directly the *amplitude* of the diffracted wave of (p/N)th order in Rowland's elementary theory of diffraction.[2] Rowland's amplitude could be actually indicated simply as $a_p(v+v_0)$ were it not for the rather immaterial fact that Rowland's variable of integration is the linear element along the grating's profile $ds=(1+\zeta'^2)^{\frac{1}{2}}dx$ instead of the linear element dx which appears in (6′). The semiquantitative arguments of the elementary theory which serve to estimate the amplitude of the light diffracted by an imperfect grating between the normal "orders" serve in the present theory to estimate the values of the coefficients a_p for a "fractional order" p/N. Through this indirect path they serve to estimate the amplitudes of the diffracted "ghost" waves.

There remains to be discussed the order of magnitude of the coefficients $c_{n'n}$ within each square submatrix, i.e., of those coefficients for which the difference $p-q$ is a multiple of N. As indicated in the preceding qualitative discussion, even these coefficients will be small unless $p-q$ is a *small* multiple of N. The same holds for the quantities $d_{n'}$, unless the index p is a *small* multiple of N. The elementary theory of accidental errors of ruling can be applied in this connection to the evaluation of the coefficients a_p for which p is a multiple of N. The well-known result is of the type

$$a_p \sim \exp[-(p\epsilon/N)^2], \tag{12}$$

where ϵ is a measure of the size of accidental errors. For example, if these errors are merely due to improper positioning of the grooves, ϵ is $\sqrt{2}\pi$ times the ratio of the r.m.s. deviation of the grooves from their theoretical position to the mean spacing of the grooves.[14] It follows then that coefficients a_p, for which the ratio p/N is sufficiently large, can be taken as equal to zero in the first approximation. The remaining elements of the submatrices of coefficients $c_{n'n}$ will therefore be concentrated about their diagonal line and only a few of the quantities $d_{n'}$ will be different from zero in the first approximation. The infinite system of Eqs. (11) is then reduced for practical purposes to a finite one and hence is amenable to numerical solution.

IV. DISCUSSION

The problem of determining the intensity distribution of the light diffracted by a finite imperfect grating has been reduced to the problem of solving a system of algebraic linear equations by following a certain approximation procedure. It should be possible to solve this system numerically without much difficulty for any particular set of given data. These must include the color, polarization, and angle of incidence of the light, the optical characteristics of the grating metal, and the equation of the grating profile. It is also possible that the method of approximation proposed here is susceptible of further analytical development. One might, for example, try to

---

[14] No specific reference is given by Sparrow (see references 1 and 11) for this type of result. A paper by Lord Rayleigh (Phil. Mag. **37**, 498 (1919) and *Coll. Works*, Vol. 6, p. 627) is pertinent to this question. A proof of (12) is outlined here. The coefficient a_p can be written in the form

$$a_p(v)=(1/N)\sum_{m=0}^{N-1}(N/L)\int_{mL/N}^{(m+1)L/N}\exp[2\pi i(v\zeta(x)-px/L)]dx$$
$$=(1/N)\sum_{m=0}b_{pm}(v).$$

If the ruling is perfect except that each groove may be displaced from its theoretical position by a small fraction δ of the spacing L/N and if p is a multiple of N,

$$b_{pm}=(a_p)_{\text{perf}}\exp(-2\pi ip\delta_m/N),$$

where $(a_p)_{\text{perf}}$ is the value of a_p in the absence of groove displacements. Assuming, finally, that all displacements δ_m are distributed as random errors with a r.m.s. value σ, the expected, or mean, value of a_p is the product of $(a_p)_{\text{perf}}$ and of the expected value of $\exp(-2\pi ip\delta_m/N)$, namely,

$$\int_{-\infty}^{\infty}\exp[-2\pi ip\delta_m/N-\delta_m^2/2\sigma^2]d\delta_m/(2\pi)^{\frac{1}{2}}\sigma$$
$$=\exp(-2\pi^2\sigma^2p^2/N^2).$$

exploit the fact that adjacent submatrices of the matrix of coefficients $c_{n'n}$ (corresponding to successive values of the index s) are approximately equal.

The mathematical expressions of the elementary theory which are meant to represent the amplitude of the light scattered in different directions appear in Section III to determine merely its order of magnitude. This is not surprising in view of the limited significance of the elementary theory, as stressed in Section I. However, a further development of the formalism presented in Section III might well show a closer connection between the results of the elementary theory and those of a more elaborate one.

The earlier explanation of the anomalous diffraction[7] fits well into the present treatment. The application of the information on the matrix $c_{n'n}$ obtained in this paper to the discussion in the appendix of reference 7 supports the argument that the convergence of Rayleigh's approximation is not essential for the explanation of the anomalies. The frequent disappearance of the anomalies upon light rubbing of the grating acquires new significance in the light of the results obtained here concerning the effect of accidental irregularities upon the matrix c.

I wish to thank Dr. R. D. Huntoon for calling my attention to this problem and for numerous useful discussions and suggestions.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 11 NOVEMBER, 194

Transmission Filters for the Ultraviolet*

M. KASHA
Department of Chemistry, University of California, Berkeley, California

(Received June 7, 1948)

The problems involved in the development of ultraviolet transmission filters are discussed and criteria of stability, efficiency, and ease of preparation are applied to the materials suggested by previous workers. By the selection of the most suitable previous filters and the introduction of several new filters, a set of band pass filters for the 2400–3800A region of the ultraviolet is developed. These have a half-width of 200A and a maximum transmission of 30 percent, on the average. Quantitative spectrophotometric data are given for all components and filter combinations used. Suggestions are made for the further development of these tools.

I. INTRODUCTION

CHEMICAL solution filters offer the best means of procuring high intensity monochromatic ultraviolet sources, when used in conjunction with suitable arcs. In this paper we shall review the materials proposed by earlier workers,[1–4] and shall extend the range and efficiency of their filter combinations by the introduction of several new filters. Particular emphasis will be placed on stability and ease of preparation of the filters, and on complete coverage of the ultraviolet range from 2400 to 3800A. Quantitative spectrophotometric data on all of the filter components and the final combinations are given, since these are lacking in most of the older work. The present set of ultraviolet filter combinations supplements the efficient visible filter set developed by Winther.[5]

In Sections II, III, and IV the general problems of filter selection are outlined, and the criteria of each filter class are illustrated by numerous examples. The spectrophotometric data are given as percent transmission in order to permit a quick appraisal of the utility of a particular filter at its optimum concentration. It is

* This work was done under a contract between the Office of Naval Research and the University of California.

[1] H. L. J. Bäckström, Naturwiss. **21**, 251 (1933).

[2] H. Alterthum and M. Reger, Chem. Fabrik **6**, 283 (1933); cf. R. Rompe, Acta Phys. Polonica **5**, 151 (1936).

[3] E. J. Bowen, J. Chem. Soc. 76 (1935). Cf. Proc. Roy. Soc. **A154**, 349 (1936). See also, E. J. Bowen, *The Chemical Aspects of Light*, 2nd Ed. (Clarendon Press, Oxford, 1946). Appendices II and III.

[4] K. S. Gibson, "Spectral filters," *International Critical Tables*, Vol. 5, p. 271 (1929).

[5] Chr. Winther, Zeits. f. Elektrochemie **43**, 691 (1937).

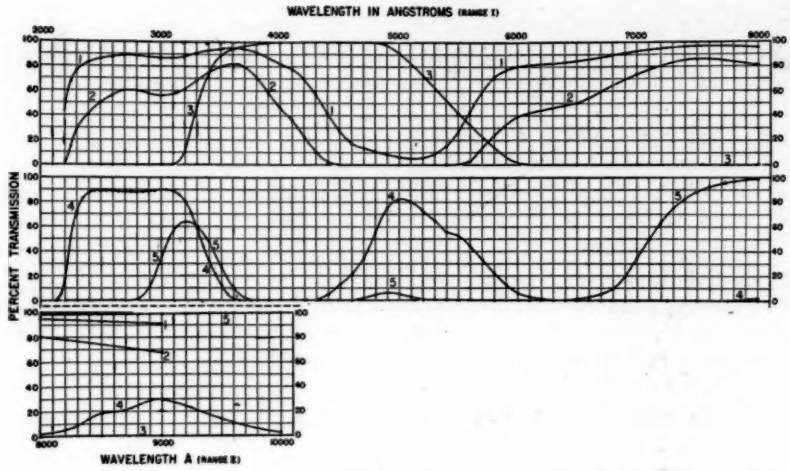


FIG. 1. Light transmission of aqueous solutions of inorganic salts, 2000 to 10,000A. (1) $CoSO_4 \cdot 7H_2O$, 75 g/l, 1-cm path. (2) $CoSO_4 \cdot 7H_2O$, 300 g/l, 1 cm. (3) $CuSO_4 \cdot 5H_2O$, 100 g/l, 5 cm. (4) $NiSO_4 \cdot 6H_2O$, 500 g/l, 1 cm. (5) $KCr(SO_4)_2 \cdot 12H_2O$, 150 g/l, 1 cm.

hoped that the presentation made will lead to the enhancement of the future development of these important tools.

II. GENERAL ULTRAVIOLET TRANSMITTING FILTERS

The first problem in the development of an ultraviolet monochromatizing filter is to find a combination which will absorb all visible and infra-red light and transmit only the ultraviolet. R. W. Wood[6] recommended a solution of *p*-nitrosodimethylaniline as a filter approaching these requirements, but this substance is not sufficiently stable to permit prolonged use. However, in several inorganic salts and glasses the requirements are met satisfactorily.

$NiSO_4$—$CoSO_4$

It has long been known that nickelous sulfate has good ultraviolet transmission characteristics. This may be seen in curve 4, Fig. 1, in which a remarkable transmission from 2300 to 3300A is evident. However, nickel sulfate has high transmission peaks at 5000 and 9000A as well, so that these must be removed by some other component. Cobaltous sulfate, which also transmits in the ultraviolet, although not as uniformly as nickel (especially in the valuable 2300–2500A region) (see curves 1 and 2, Fig. 1), does absorb well in the visible and infra-red transmission regions of nickel sulfate. From the nature of the ultraviolet transmission of cobalt sulfate, it was concluded that the ratio of cobalt to nickel should be kept to the minimum permitted for the desired absorption in the visible and infra-red.

It was found that a 5.0 cm optical path of an aqueous solution containing 240 g/l of $NiSO_4 \cdot 6H_2O$ and 45.0 g/l of $CoSO_4 \cdot 7H_2O$, would transmit approximately 50 percent of the incident light between 2400 and 3200A, as shown in Fig. 2, and no other light between 2000 and 10,000A except in two localized regions of the spectrum: a hump of 1.5 percent at 5700A, and 3.9 percent at 9000A. The latter would undoubtedly be totally absorbed by the 5 cm of water[7] (the experimental procedure gave the *selective* absorption of the dissolved salt). Bäckström[1] recommended a ratio of 276/84.3 of Ni/Co, quoted in g/l of the above compounds respectively (3 cm layer of $1.75M$ $NiSO_4$ and $0.5M$ $CoSO_4$) for the same path length of 5 cm. The larger amount of cobalt would remove the transmitted visible light more completely, although at an unnecessary sacrifice to the transmission at 2500A. Bowen[3] used 5 cm of 290 g/l $NiSO_4 \cdot 6H_2O$, 83.0 g/l $CoSO_4 \cdot 7H_2O$. The Co-Ni

[6] R. W. Wood, Phil. Mag. **5**, 257 (1903).

[7] Cf. Fig. 1, T. Dreisch, and W. Trommer, Zeits. f. physik. Chemie **B37**, 37 (1937).

solution preserves its initial transmission remarkably well even after tens of hours of exposure to the direct light of the *AH*-6 high pressure quartz mercury arc.

The visible transmission of the Ni-Co filter may be reduced to a negligible value by the use of 1.0 mm of Corning Glass c.s. 7-54 (9863) without seriously affecting the 2500A transmission (see below, however).

$NiSO_4$—Red Purple Corex A Glass

The Corning Glass Works[8] "red purple corex A" filter c.s. 7-54 (Old no. 9863 or 986) (3 mm) is similar to cobalt sulfate in its ultraviolet and visible transmission characteristics and may be used conveniently in its place. The visible transmission with a 3 mm Corning 7-54 filter and a 5.0 cm path of 200 g/l $NiSO_4 \cdot 6H_2O$ is negligible. Although at 9000A there is somewhat more transmission than with the Ni-Co combination, the 5 cm layer of water would remove this.[7] A new Corning 7-54 filter transmits "a minimum of 40 percent" of the incident light at 2500A; whereas an old filter in our laboratory, which had undergone long exposure to the light of the quartz mercury arc, had its 2500A transmission reduced to 20 percent (cf. Fig. 2). This phototropism may be partly reversed by baking the filter in an oven at 450°C, according to the old Corning Glass filter book.[9] It is recommended that, in view of the results demonstrated by Fig. 2, old red purple corex filters be relegated to work in the 2800–3700A region (unchanged by light) and new filters be reserved for 2300–2600A work, in order to achieve the best high frequency transmission.

$CrK(SO_4)_2$

Chrome alum has a good broad transmission band at $\lambda_{max} = 3200$A, as shown by curve 5, Fig. 1. There is a little transmission in the visible, a broad region with 6-percent transmission at $\lambda_{max} = 4900$A. This could be removed by $CoSO_4$ (curve 2, Fig. 1), which would still transmit most of the light at 3200A. However, chrome alum and cobalt sulfate also transmit most of the infra-red from 6500 to 10,000A. This could be removed by $CuSO_4$.

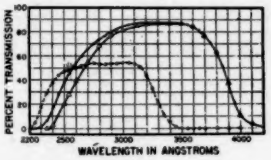


FIG. 2. General ultraviolet-transmitting filters. *Dotted curve:* ultraviolet transmission of a 5.0-cm optical path of an aqueous solution containing 240 g/l of $NiSO_4 \cdot 6H_2O$ and 45.0 g/l of $CoSO_4 \cdot 7H_2O$. *Solid curves:* ultraviolet transmission of Corning Glass filter c.s. 7-54 (9863 or 986, "red purple corex A"), 3.0 mm; *circles:* new filter; *triangles:* old filter (after exposure to the light of the AH6 high pressure mercury arc for over 300 hours).

$CuSO_4$

Cupric sulfate is the most useful general filter for removing the infra-red, when only the near ultraviolet and visible are to be isolated. Curve 3, Fig. 1, shows the transmission of a 5.0 cm path of a 100 g/l aqueous solution of $CuSO_4 \cdot 5H_2O$. Its stability in intense mercury arc is good.

III. REGION-SELECTING FILTERS

Using the general UV-transmitting filter combinations described in Section II, it is next necessary to narrow down the range of transmission by finding filters which will transmit but a part of these broad regions. This is a difficult problem inherently and the one in which the least work has been done. There are many substances which begin to absorb at some wave-length and continue to absorb more or less strongly at all shorter wave-lengths (see Section IV). On the other hand, very seldom does a substance have a strong absorption band separated from other higher energy absorption bands by a region of *low* absorption. However, a search of some pub-

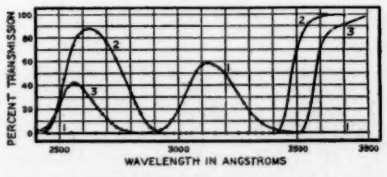


FIG. 3. Region-selecting filters. (1) K_2CrO_4, 0.200 g/l water, 1-cm path. (2) (cn), iodide in water, 20 mg/100 ml, 1 cm (a simple cyclic cyanine-like substance, c.f. reference 10). (3) 1,4-diphenylbutadiene (DPBD), 4.24 mg/100 ml ethyl ether, 1 cm.

[8] *Glass Color Filters* (Corning Glass Works, Corning, New York, 1948), New Edition.
[9] *Glass Color Filters* (Corning Glass Works, Corning, New York, 1942) (Old Edition, Form C-247-4-42).

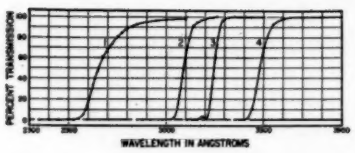


FIG. 4. High frequency cut-off filters. (1) Carbon tetrachloride, 0.5-cm path. (2) Potassium acid phthalate, 5.0 g/l water, 1 cm. (3) Naphthalene in isoöctane, 12.8 g/l, 1 cm. (4) (cn), iodide in water, 20 mg/100 ml, 1 cm (c.f. reference 10, and curve 2, Fig. 3).

lished spectra indicated that a number of suitable "region selecting" or isolating filters are available. Three useful ones are documented here.

K_2CrO_4

Potassium chromate, as is well known, has a good transmission maximum at 3130A, exactly the wave-length of a group of strong mercury lines. Curve 1, Fig. 3 shows the transmission spectrum of a 1-cm path of a water solution containing 0.200 g/l of K_2CrO_4. If neutral potassium chromate is not available, it would be preferable to dissolve the salt in 0.05*N* sodium hydroxide in order to prevent the formation of the dichromate, which absorbs light differently. It is important to recognize the fact that the solution of potassium chromate transmits also from 2150 to 2450A with about 15 percent transmission at $\lambda_{max} = 2300$A. If a smaller transmission of the 3341A Hg line is required, 0.300 g/l may be used (45 percent at 3130, and 3 percent at 3340A).

A Simple Cyclic Cyanine-Like "Dye"

This has a complicated chemical name,[10] but we shall designate it by the code (cn). It serves as an exceedingly good ultraviolet transmitting filter. It has an exceptional chemical and photochemical[11] stability. Because of its intense absorption at about 3200A, and its electronic simplicity, it has a very strong transmission peak at about 2600A. Thus, a water solution containing 20 mg/100 ml of the iodide of this dye transmits from 2450 to 2900A, with about 90 percent transmission at λ_{max} (curve 2, Fig. 3). The usefulness of this filter depends on the fact that it absorbs very strongly in the 3000 to 3400A transmission region of nickel sulfate.

1,4-Diphenylbutadiene

The absorption spectrum of this substance,[12] which for our purpose may be designated by DPBD, shows a pronounced transmission peak at about 2600A. A solution containing 4.24 mg of DPBD (recrystallized) in 100 ml ethyl ether (reagent grade, redistilled) shows the useful transmission peak given by curve 3, Fig. 3. This solution has a transmission of over 40 percent at 2550A close to the mercury resonance line, and no transmission from 2900 to 3500A. The limiting factor in the 2500A transmission is the inefficiency of $CoSO_4$ or red purple corex at this wave-length. Preliminary experience with the DPBD filter in the beam of a Hanovia (low pressure) quartz mercury arc indicates good stability.

Other Region-Selecting Filters

If other region-selecting filters are desired, a perusal of the standard compilations of absorption spectra (*International Critical Tables*, Vol. **5**; Landolt-Börnstein *Physikalisch-Chemische Tabellen;* consult indices) of inorganic and organic substances will be found useful. In general, (*a*) the substance must be chemically stable (e.g., not hydroquinone: although this has a good transmission band at 2500A, it is readily oxidized by the oxygen of the air); (*b*) the intensity of the absorption on the long wave-length side of the transmission band must be at least 1000 times stronger than at the "transmission band" (e.g., not KNO_3 and other nitrates, in which the absorption at 2950 is only about 10 times the absorption at 2600A), and (*c*) the substance should be moderately stable in light (e.g., not such substances as sodium cinnamate and isoeugenol, which absorb light in a useful manner, but are rapidly photolyzed).

Several substances, untested here, which seem to have good ultraviolet transmission characteristics, are (1) methyl ester of terephthalic acid,

[10] Cation *X*, named 2,7-dimethyl-3,6-diazacyclohepta-1,6-diene iodide. G. Schwarzenbach and K. Lutz, Helv. Chim. Acta **23**, 1139 (1940). The authors give directions for preparation of the perchlorate; the iodide was used in our work.

[11] In our use of the filter, little measurable change in transmission occurred upon prolonged exposure to the high pressure quartz mercury arc (the light first being passed through nickel sulfate and red purple corex).

[12] Landolt-Börnstein, *Physikalisch-Chemische Tabellen* (Julius Springer, Berlin, 1935). Curve 141, p. 1379, Volume 111b.

transmits at <2100A (L. B. tables,[12] Vol. **111b**, p. 1414, Fig. 246), (2) resorcinol dimethylether, 2450A (L. B. Vol. **111b**, p. 1417, Fig. 256), (3) naphthacene, <2500A (L. B. Vol. **111b**, p. 1391, Fig. 176).

IV. HIGH FREQUENCY CUT-OFF FILTERS

In order to limit the broadness of the ultraviolet band transmitted by the "general UV-transmitting" followed by the "region-selecting" filter, a high frequency cut-off filter may be used. These are very common, and the choice depends on the requirements of (*a*) the cut-off wavelength, (*b*) the sharpness of the cut-off, and (*c*) the stability of the filter. The glass filters[8] include several which have ultraviolet cut-off limits; but in general they are too gradual in their transmission characteristics for our purpose. Chemical filters often have very sharp ultraviolet cut-off limits, and many investigators have made use of and described them. Four useful ones are described here, two being new.

Carbon Tetrachloride

Curve 1, Fig. 4, gives the transmission curve of a 0.5-cm optical path of liquid CCl_4. It is useful for eliminating the 2536A line of Hg. If only a 1-cm quartz cell is available, the CCl_4 may be diluted with an inert, non-absorbing liquid such as methylcyclohexane. This filter has been used by Bowen.[3] It is moderately stable in light. A sharpter cut-off would be more desirable.

Potassium Acid Phthalate

Curve 2, Fig. 4, gives the transmission of a 1-cm path of an aqueous solution containing 5.0 g/l of $KHC_8H_4O_4$ (0.0245*M*). This filter was described by Bäckström[1] and has been discussed recently by Hunt and Davis.[13] It isolates the 3130A group of Hg lines from those transmitted around 3000A by the K_2CrO_4 region-selecting filter. It is slightly photochemically unstable and requires occasional renewal. In several cases in this laboratory, bacterial mold has been observed to grow in the biphthalate solution. This might be eliminated by preliminary sterilization of the water and stoppering of the solution. A solution containing a moldy growth obviously should not be used.

[13] R. E. Hunt and W. Davis, Jr., J. Am. Chem. Soc. **69**, 1415 (1947).

Naphthalene

Curve 3, Fig. 4, gives the transmission curve of a 1-cm path of an isoöctane solution of naphthalene, containing 12.8 g/l. This filter serves to isolate the 3340A Hg line by completely removing all light of shorter wave-length than 3160A, in combination with other filter components. It should be pointed out that the *fluorescence* of naphthalene (3200 to 3900A) must be taken into account in some applications of this filter; however, the quantum yield is about 0.2 in dilute solutions, and is very probably much less in a solution of the concentration suggested above. The effect of the fluorescence may be reduced still further by placing the cell last in the combination sequence, so that it absorbs only the last part of the unwanted light (see Section V).

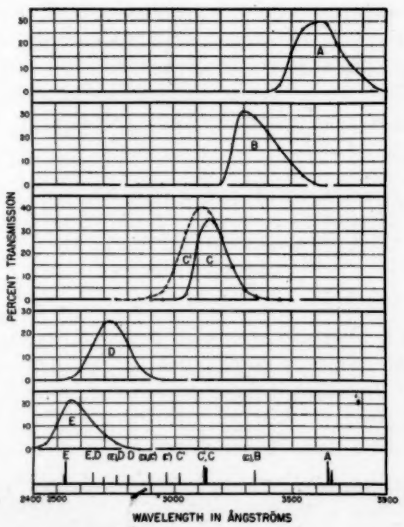


FIG. 5. Band-pass filter combinations for the ultraviolet. The combinations consist of the following components, the numbers corresponding to those given in Table I. $A=(1+6+9)$. $B=(2+7+10)$. $C=(3+8+11+12)$. $C'=(3+8+11)$. $D=(4+8+9+13)$. $E=(5+14)$. Just above the wave-length scale is given a schematic representation of the persistent lines of the low pressure mercury arc. The letters just above these lines correspond to the combinations which will transmit the respective lines. See Section V of the text for specific details of the use of these filters and notes on "filter defects."

TABLE I. Filter—Combination Components.

Component	Substance	Solution	Optical path
1.	$CuSO_4 \cdot 5H_2O$	100 g/l water	5 cm
2.	$NiSO_4 \cdot 6H_2O$	100 g/l water	5 cm
3.	$NiSO_4 \cdot 6H_2O$	200 g/l water	5 cm
4.	$NiSO_4 \cdot 6H_2O$	300 g/l water	5 cm
5.	$NiSO_4 \cdot 6H_2O$	240 g/l water	
and	$CoSO_4 \cdot 7H_2O$	45.0 g/l water	5 cm
Corning filters			
6.	C. S. 7-37 (5860 or 586)		5 mm
7.	C. S. 7-51 (5970 or 597)		4 mm
8.	C. S. 7-54 (9863 or 986)		3 mm
9.	(cn),* (iodide)	20 mg/100 ml water	1 cm
10.	Naphthalene	12.8 g/l isoöctane	1 cm
11.	K_2CrO_4	0.200 g/l water	1 cm
12.	$KHC_8H_4O_4$	5.0 g/l water	1 cm
13.	CCl_4	Pure liquid	0.5 cm
14.	1,4-Diphenylbutadiene	4.24 mg/100 ml ether	1 cm

* See reference 10.

The Cyanine Dye (cn)

Curve 4, Fig. 4, gives the transmission of the same solution we used in Section III, but depicting only the long wave-length transmission limit. This is for 1 cm of a water solution containing 20 mg/100 ml of iodide of the cyclic "cyanine dye." Its sharp ultraviolet cut-off makes it exceptionally well suited to the isolation of the 3650A Hg line group. Other properties of this substance have been described in Section III. The sharpness of this chemical filter makes it much superior to alternative commercial glass filters.

V. BAND PASS FILTER COMBINATIONS FOR TRANSMISSION OF 200A REGIONS OF THE ULTRAVIOLET BETWEEN 2400 AND 3800A

In Fig. 5 the transmission curves are given for the set of monochromatizing filters. The component numbers correspond to those given in Table I, which lists the composition of each component.

In using these combinations, the recommended order is (1) 5-cm quartz cell of the inorganic salt solution, (2) the glass filter, (3) the 1-cm quartz cells of the remaining components. Placing the 5-cm cell first cools the light beam, and placing the 1-cm cells last protects these (generally) less stable solutions from unnecessary exposure.

A filter combination consisting of several components—each of which has several transmission regions in addition to its principle transmission band—often has subsidiary low transmissions, which must be regarded as *filter defects*. Thus, Combination E transmits 1.5 percent around 5700A, and as mentioned under "$NiSO_4 - CoSO_4$" in Section II, this may be eliminated by the use of 1.0 mm of Corning Glass c.s. 7-54 (Filter 9863). Combination E with 1.0 mm of 9863, and Combinations A, B, and C are devoid of significant transmission other than the principle region described by Fig. 5. Combination D, however, does show a characteristic defect, consisting of a rather narrow band of transmission at about 3470A. The filter defects were sought by long exposure experiments using the AH6 high pressure mercury arc and a medium Hilger quartz spectrograph, and represent less than 0.1-percent transmission. The presence of about 5 cm of water in the optical path insures[7] the absence of light transmission between 9000 and 20,000A.

A large amount of variation may be introduced into the properties of the filter combinations by the use of various ultraviolet sources. In general it has been found that the AH6 high pressure mercury arc gives good emission intensities at all wave-lengths in the ultraviolet range considered, except at 2500–2600A. This latter region is devoid of emission as a result of the quenching of the Hg resonance line. Thus, the AH6 arc has given satisfactory results with combinations A, B, C, and D, while a Hanovia medium low pressure quartz mercury arc has been found useful for the 2400–2500A region (combination E).

The chemicals used in the filters described in this paper were all of C. P. reagent grade. The absorption curves were obtained by means of a Beckman Quartz Spectrophotometer, Model DU.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 11 NOVEMBER, 1948

A Differential Adjustment Method of Refining Optical Systems

WALDEMAR M. STEMPEL
Physics Department, Stevens Institute of Technology Hoboken, New Jersey
(Received July 12, 1948)

It is assumed that a preliminary design of an optical system has been completed and the characteristic rays traced through it to determine the aberrations which afflict it. The present article then starts with these data and offers a set of reduction formulae. These formulae are applied, surface by surface, from the image space forward until any particular parameter is reached. This is done for every parameter in the system. These results are finally substituted in differentiated aberration formulae to give a coefficient which, when multiplied by the change in the parameter, gives the change in the aberration in the final image space. When these coefficients are calculated for all parameters and all aberrations and recorded in a table, say, with the aberrations across the top and the parameters down the side, the designer has a complete picture of the performance of that optical system. The table then enables him to quickly adjust all parameters so as to obtain the best results possible from the set-up as given.

IN an earlier article, "Empirical approach to lens design,"[1] the author offered a group of slide rule equations for quickly computing changes in vergence angles and intersecting lengths of a refracted ray caused by slight changes in the radius of curvature of the refracting surface or slight changes in the index of refraction.

At that time, attempts made to apply these equations, in one step, to an entire series of refracting surfaces making up the lens system, resulted in seemingly hopeless complications, and so the equations were considered useful only in passing, surface by surface, from an object space to the corresponding image space. This was a considerable time saver over retracing the rays, especially since the equations dealt with the changes only and not with the principal quantities, which fact made the precision of the ordinary slide rule quite adequate.

It was, however, fully realized that, if these operations could be simplified for the entire system, the lens designer would have an additional powerful tool for rapidly making corrections vital to producing highly corrected lens systems. The author believes he has now accomplished this task and herein offers his findings.

This work has been in preparation since the publication of the author's first article, previously referred to. Two years later, and after the method here offered was completed, except for some time consuming details, there appeared a series of articles on a method for accomplishing the same thing by A. L. M'Auley and F. D. Cruickshank of Tasmania University.[2] They very kindly sent me reprints, and the author used their method to check his own, and found that the two methods check each other quite well. The Cruickshank method is, however, less direct and less automatic, and calls for the determination of parameters not too commonly used by the designer. The method here described calls for only such data as are already available in the original ray tracings. Both methods are subject to errors which creep in (as a result of the calculus) when the parameters are larger than infinitesimals, for which corrections are readily obtained when the need arises. A general correction formula is given in this article.

Neither method (mine nor the Cruickshank) checked with ray tracings. The publication of this work was therefore delayed because of the difficulty encountered in attempting to check against ray tracings. These difficulties are described later in this article.

It is, of course, common knowledge that, while certain changes anywhere in an optical system will have definite effects upon certain aberrations, for instance, bending an element will change the spherical aberration of the system, these same changes will also affect, generally to a lesser degree, all other aberrations. Therefore, it becomes

[1] W. M. Stempel, J. Opt. Soc. of Am. **33**, 278 (1943).

[2] A. L. M'Auley, Proc. Phys. Soc. **LVII** 302, 350, 419, 435 (1945). F. D. Cruickshank, J. Opt. Soc. Am. **36**, 13, 103 (1946).

TABLE I.

	Marginal ray		Paraxial ray	
	Curved surface	Plane surface	Curved surface	Plane surface
$M=\partial U'/\partial U$	[b]$1+G\cot U$	$N\cos I/N'\cos I'$	[b]$1+(l-r)(N'-N)/N'r$	N/N'
$O=\partial U'/\partial L$	$G/(L-r)$	0	0	0
$P=\partial U'/\partial r$	$-LG/r(L-r)$	0	0	0
$Q=\partial U'/\partial N/N'$	[a]$-\sin I/\cos I'$	[a]$-\sin I/\cos I'=\sin U/\cos U'$	0	0
$R=\partial L'/\partial U=\partial(L'-r)/\partial U$	[b]$E(\cot U-M\cot U')$	[b]$(L/\tan U')(\sec^2U-\sec^2U')$	[b]0	[b]0
$T=\partial L'/\partial L=\partial(L'-r)/\partial L$	$F-EO\cot U'$	$\tan U/\tan U'=N'\cos U'/N\cos U=N'\cos I'/N\cos I$	$(l'/l)\{(N/N')-(l-r)(N'-N)/N'r\}$	$+N'/N$
$V=\partial L'/\partial r=[\partial(L'-r)/\partial r]+1$	$1-F-EP\cot U'$	0	$1-(Nl'/N'l)+[l'(l'-r)/r^2]\,(N'-N)/N'$	0
$W=\partial L'/\partial N/N'=\partial(L'-r)/\partial N/N'$	[a]$E\{(N'/N)-Q\cot U'\}$	[a]$\{N'L/N\cos U\}\{\sin U'\sin I/\cos I'-(N'/N)\cos U'\}$	[a]$(N'/N)(l'/r)(l'-r)$	[a]$-L[N'/N]^2$

[a] This quantity is not to be used alone, since, when the glass is changed, there will be changes at both surfaces of the element, therefore the entire element must be considered as a unit, and we have,

$$Q_{12}=\underset{\text{(entire element)}}{\partial U_2'/\partial N_1/N_1'}=Q_1+\left[\left(\frac{N_1}{N_1'}\cos I_2'-\cos I_2\right)\Big/\frac{N_1}{N_1'}\cos I_2'\cos I_2\right]\left[Q_1\sin I_2\cot U_2+W_1\frac{\sin U_2}{r_2}\right]+\left[\frac{N_1}{N_1'}\right]^2\frac{\sin I_2}{\cos I_2'}.$$

$$W_{12}=\underset{\text{(entire element)}}{\partial L_2'/\partial N_1/N_1'}=(L_2'-r_2)\left(W_1\frac{\sin U_2}{r_2\sin I_2}+Q_1\cot U_2-Q_{12}\cot U_2'-\frac{N_1'}{N_1}\right)$$

For the paraxial ray these become,

$$Q_{12}\text{ par}=0,$$

$$W_{12}\text{ par}=+\frac{W_1(l_2'-r_2)}{l_2-r_2}-\frac{N_1'}{N_1}\frac{l_1'-r_1}{r_1}(l_2'-r_2)+\frac{l_1'-r_1}{r_1}\frac{(l_2'-r_2)^2}{l_2-r_2}+\frac{N_1-N_1'}{N_1}\frac{l_1'-r_1}{r_1}\frac{(l_2'-r_2)^2}{r_2}-\frac{N_1-N_1'}{N_1'r_2}\frac{(l_2'-r_2)^2}{l_2-r_2}W_1-\frac{N_1'}{N_1}\frac{(l_2'-r_2)^2}{r_2}-\frac{N_1'}{N_1}(l_2'-r_2).$$

[b] For the case where the incident beam is parallel to the axis of the system, always the case for the first surface when designing photographic lenses, and, under the conditions associated with partial derivatives, the $\partial U'/\partial U$ is infinite and the $\partial L'/\partial U$ indeterminate. In taking these partial derivatives the beam is rotated about the point where the beam cuts the axis of the system. For a beam that is parallel to the axis of the system, the results obtained in this way have no useful interpretation. We are, in this case, more interested in turning the beam about the point where it cuts the refracting surface, so that the height of this point above the axis is constant. L is then a variable and its differential coefficient must be retained. If the differentiation is carried out in this manner, the partials under consideration take the following form.

$\partial U'/\partial U=+(N/N')(\cos I/\cos I')$ for the marginal ray, N/N' for the paraxial ray,

$\partial L'/\partial U=-(N/N')(\cos I/\cos I')[r\cos I'+(l'-r)\cos U']/\sin U'$ for the marginal ray,

$=-(N/N')(L'/u')$ for the nominal paraxial ray.

especially valuable information for the optical designer to have a set of computing equations which will enable him to determine quickly and easily what effects will be produced on each of the various aberrations by a change in any of the parameters of the system. This is the purpose of the equations to be described in this article.

The essential equations, given in the article previously referred to, will be repeated here, as they form the starting point for the work given in this article.

Notations to be Used

U = incident vergence angle U' = refracted vergence angle
I = angle of incidence I' = angle of refraction
L = axial object distance L' = axial image distance

Both L and L' are measured from the pole of the refracting surface, the point where the axis cuts the refracting surface.

N = index for the object space N' = index for the image space

All the above refer to any single refracting surface.

N_0^0 = index for the object space of an entire optical system.
N'' = and all other double primed letters refer to the final image space of the complete optical system.
F'' = the focal length of the entire optical system.
U_{pr0} = Incident angle of the principal ray for the entire optical system.

Other zero subscripts will refer to the entire system.

r = Radius of curvature of the refracting surface.

$$E=r\frac{\sin I'}{\sin U'}.\quad F=\frac{N\sin U}{N'\sin U'}.\quad G=\frac{\sin(I-I')}{\cos I\cos I'}.$$

All lower case letters are for the same quantities as above except that they refer to the paraxial region of the system.

With the above notation Table I, taken from the previous article will be intelligible.

DEDUCTION OF THE EQUATIONS

We recall the general differential equations,

$$dU'=\frac{\partial U'}{\partial U}dU+\frac{\partial U'}{\partial L}dL+\frac{\partial U'}{\partial r}dr+\frac{\partial U'}{\partial N/N'}d\frac{N}{N'},\quad(1)$$

and

$$dL'=\frac{\partial L'}{\partial U}dU+\frac{\partial L'}{\partial L}dL+\frac{\partial L'}{\partial r}dr+\frac{\partial L'}{\partial N/N'}d\frac{N}{N'}.\quad(2)$$

These equations are applied to that surface in an optical system where it is proposed to make changes. That surface constitutes the first of a series extending to the end of the system in the direction in which the ray tracing was executed.

For convenience they will be written,

$$dU_1' = M_1 dU_1 + O_1 dL_1 + P_1 dr_1 + Q_1 d(N/N'), \quad (3)$$

$$dL_1' = R_1 dU_1 + T_1 dL_1 + V_1 dr_1 + W_1 d(N/N'). \quad (4)$$

If the surface is the last surface of the optical system, Eqs. (1) and (2) give all the information about U_1'' and L_1'' desired. (Note. It must be remembered, however, that a change in the index of refraction of this element is accompanied by a change in U and L at the first refracting surface of that element of the system.)

If changes in parameters are to be made at the next to the last surface, there are different courses dictated according to the demands of the problem.

If, as is usually the case, the radius of curvature of this surface is to be changed in such a way as to keep the focal length of the element constant, then the second radius of curvature must also be changed so that,

$$dr_2 = \left(\frac{r_2}{r_1}\right)^2 dr_1. \quad (5)$$

If the object and the image of the element are in the same medium, it might furthermore be a little more convenient if $d(N/N')$ be changed to dN' by writing

$$d\frac{N_1}{N_1'} = -\frac{N_1}{N_1'^2} dN_1',$$

and (6)

$$d\frac{N_2}{N_2'} = d\frac{N_1'}{N_2'} = \frac{1}{N_2'} dN_1'.$$

When only the last two surfaces are under consideration, the differential equation for the second of the two surfaces and therefore the last surface of the system is

$$dU_2' = \frac{\partial U_2'}{\partial U_2} dU_2 + \frac{\partial U_2'}{\partial L_2} dL_2 + \frac{\partial U_2'}{\partial r_2} dr_2 + \frac{\partial U_2'}{\partial N_2/N_2'} d\frac{N_2}{N_2'},$$

or

$$= M_2 dU_2 + O_2 dL_2 + \left(\frac{r_2}{r_1}\right)^2 P_2 dr_1 + \frac{Q_2}{N_2'} dN_1',$$

but since

$$dU_2 = dU_1' = M_1 dU_1 + O_1 dL_1 + P_1 dr_1 + Q_1 d\frac{N_1}{N_1'},$$

and

$$dL_2 = dL_1' = R_1 dU_1 + T_1 dL_1 + V_1 dr_1 + W_1 d\frac{N_1}{N_1'},$$

$$dU_2' = M_2\left(M_1 dU_1 + O_1 dL_1 + P_1 dr_1 + Q_1 d\frac{N_1}{N_2}\right) + O_2\left(R_1 dU_1 + T_1 dL_1 + V_1 dr_1 + W_1 d\frac{N_1}{N_1'}\right) + \left(\frac{r_2}{r_1}\right)^2 P_2 dr_1 + \frac{Q_2}{N_2'} dN_1'.$$

Collecting terms,

$$dU_2' = (M_2 M_1 + O_2 R_1) dU_1 + (M_2 O_1 + O_2 T_1) dL_1 + \left(M_2 P_1 + O_2 V_1 + \left(\frac{r_2}{r_1}\right)^2 P_2\right) dr_1 + \left(-\frac{N_1}{N_1'^2} M_2 Q_1 - \frac{N_1}{N_1'^2} O_2 W_1 + \frac{Q_2}{N_2'}\right) dN_1'.$$

Similarly,

$$dL_2' = (R_2 M_1 + T_2 R_1) dU_1 + (R_2 O_1 + T_2 T_1) dL_1 + \left(R_2 P_1 + T_2 V_1 + \left(\frac{r_2}{r_1}\right)^2 V_2\right) dr_1 + \left(-\frac{N_1}{N_1'^2} R_2 Q_1 - \frac{N_1}{N_1'^2} T_2 W_1 + \frac{W_2}{N_2'}\right) dN_1'.$$

These last equations may be condensed into

$$dU_2' = H_2^U dU_1 + H_2^L dL_1 + H_2^r dr_1 + H_2^N dN_1', \quad (7)$$

$$dL_2' = K_2^U dU_1 + K_2^L dL_1 + K_2^r dr_1 + K_2^N dN_1', \quad (8)$$

where

$$H_2^U = M_2 M_1 + O_2 R_1,$$

$$H_2^L = M_2 O_1 + O_2 T_1,$$

$$H_2{}^{r} = M_2P_1+O_2V_1+\left(\frac{r_2}{r_1}\right)^2 P_2,\dagger$$

$$H_2{}^{N_1/N_1'} = M_2Q_1+O_2W_1-\left(\frac{N_1'}{N_1}\right)^2 Q_2, \tag{9}$$

$$H_2{}^{N_1'} = -N_1/N_1'^2(M_2Q_1+O_2W_1)+\frac{Q_2}{N_2'}.$$

$$K_2{}^{U} = R_2M_1+T_2R_1,$$

$$K_2{}^{L} = R_2O_1+T_2T_1,$$

$$K_2{}^{r} = R_2P_1+T_2V_1+\left(\frac{r_2}{r_1}\right)^2 V_2,\dagger\dagger \tag{10}$$

$$K_2{}^{N_1/N_1'} = R_2Q_1+T_2W_1-(N_1'/N_1)^2W_2,$$

$$K_2{}^{N_1'} = -N_1/N_1'^2(R_2Q_1+T_2W_1)-(N_1'/N_1)^2W_2.$$

Attention is called to the reason for denoting these coefficients in this way. The plan greatly simplifies the computation in the case of complex optical systems as will be explained later. But it should here be noted that the H coefficients appear in the equation for U', while the K coefficients appear in the equation for L'. The subscripts denote the surface under consideration, counted in the direction of ray tracing, while the superscripts denote the differential change for which it is the coefficient.

The calculations for the two surfaces for that lens where changes are to be made (those just discussed—Eqs. (9) and (10)) are quite complex, but, since we are dealing with differentials and not the entire quantity in which the differences occur, the precision obtainable with a slide rule is, in many cases, ample.

It must be very carefully noted that the subscripts as here used do not remain fixed to a surface. Subscript (1) is the one in which changes are being made, and subscript (2) is the next in the direction of ray tracing. As we pass from the consideration of one surface after another, going from the back (image end) to the front (object end) of the system, all subscripts are advanced by one for each surface. Though this practice is not a customary one, it greatly simplifies the formulating of this new method of lens correction.

It must be further noted that the two surfaces of any one element, with air on both sides, must be treated as a unit. If a change in the first radius is anticipated, we must decide whether the focal length of this element is to change or to be kept constant. If the latter condition is to prevail, the second radius will have to change also. If the index of refraction of the glass is to be changed, changes will occur at both surfaces. Equations (9) and (10) take care of this condition. For surfaces following the element in which changes are anticipated, the new "differential adjustment" method of lens corrections simplifies the work as follows.

If we now pass to the next surface toward the front (this puts the surface (1) previously so designated forward of the last element) and assume no changes are made in the remaining two surfaces, we then have

$$\begin{aligned} d\dot{U}_3'' &= M_3dU_3+O_3dL_3 \\ &= M_3\{M_2[M_1dU_1+O_1dL_1+P_1dr_1] \\ &\quad +O_2[R_1dU_1+T_1dL_1+V_1dr_1]\} \\ &\quad +O_3\{R_2[M_1dU_1+O_1dL_1+P_1dr_1] \\ &\qquad +T_2[R_1dU_1+T_1dL_1+V_1dr_1]\} \\ &= (M_3M_2M_1+M_3O_2R_1+O_3R_2M_1 \\ &\quad +O_3T_2R_1)dU_1+(M_3M_2O_1+M_3O_2T_1 \\ &\quad +O_3R_2O_1+O_3T_2T_1)dL_1+(M_3M_2P_1 \\ &\qquad +M_3O_2V_1+O_3R_2P_1+O_3T_2V_1)dr_1. \end{aligned}$$

Recalling that subscript (1) now refers to the third from the last surface, (2) the second from the last and (3) the last, and translating all subscripts into fixed ones based on representing the total number of surfaces in the system by (a), this equation reduces to

$$\begin{aligned} dU_a'' &= (M_{a-2}H^U{}_{a-1}+R_{a-2}H^L{}_{a-1})dU_1 \\ &\quad +(O_{a-2}H^U{}_{a-1}+T_{a-2}H^L{}_{a-1})dL_1 \\ &\quad +(P_{a-2}H^U{}_{a-1}+V_{a-2}H^L{}_{a-1})dr_1. \end{aligned} \tag{11}$$

† The third term in the right hand member is to be used if f' of the element is constant. If only the first radius is changed, and the resulting change in the focal length of that element, and consequently the focal length of the system, be permissible, then this term in P_2 disappears altogether.

†† The third term in the right hand member is to be used if f' of the element is constant. If only the first radius is changed, and the resulting change in the focal length of that element, and consequently the focal length of the system, be permissable, then this term in V_2 disappears altogether.

Similarly

$$
\begin{aligned}
dL_a'' &= R_3 dU_3 + T_3 dL_3 \\
&= R_3\{M_2[M_1 dU_1 + O_1 dL_1 + P_1 dr_1] \\
&\quad + O_2[R_1 dU_1 + T_1 dL_1 + V_1 dr_1]\} \\
&\quad + T_3\{R_2[M_1 dU_1 + O_1 dL_1 + P_1 dr_1] \\
&\qquad + T_2[R_1 dU_1 + T_1 dL_1 + V_1 dr_1]\} \\
&= (R_3 M_2 M_1 + R_3 O_2 R_1 + T_3 R_2 M_1 \\
&\quad + T_3 T_2 R_1) dU_1 + (R_3 M_2 O_1 + R_3 O_2 T_1 \\
&\quad + T_3 R_2 O_1 + T_3 T_2 T_1) dL_1 + (R_3 M_2 P_1 \\
&\quad + R_3 O_2 V_1 + T_3 R_2 P_1 + T_3 T_2 V_1) dr_1 \\
&= (M_{a-2} K^U{}_{a-1} + R_{a-2} K^L{}_{a-1}) dU_1 \\
&\quad + (O_{a-2} K^U{}_{a-1} + T_{a-2} K^L{}_{a-1}) dL_1 \\
&\qquad + (P_{a-2} K^U{}_{a-1} + V_{a-2} K^L{}_{a-1}) dr_1. \qquad (12)
\end{aligned}
$$

From these equations the "final image space" coefficients for varying parameters in the third from the last surface are obtained by inspection as

$$
\begin{aligned}
H^U{}_{a-2} &= M_{a-2} H^U{}_{a-1} + R_{a-2} H^L{}_{a-1}, \\
H^L{}_{a-2} &= O_{a-2} H^U{}_{a-1} + T_{a-2} H^L{}_{a-1}, \\
H^r{}_{a-2} &= P_{a-2} H^U{}_{a-1} + V_{a-2} H^L{}_{a-1}, \\
K^U{}_{a-2} &= M_{a-2} K^U{}_{a-1} + R_{a-2} K^L{}_{a-1}, \\
K^L{}_{a-2} &= O_{a-2} K^U{}_{a-1} + T_{a-2} K^L{}_{a-1}, \\
K^r{}_{a-2} &= P_{a-2} K^U{}_{a-1} + V_{a-2} K^L{}_{a-1}.
\end{aligned}
$$

The co-efficients for a change of index of refraction is here again omitted as the entire element must be taken as a unit.

Continuing in the above fashion, the coefficients may be determined for any number of surfaces in a system. They will here not be worked out, but for convenient reference the coefficients for the last six surfaces will be collected at this point.

Last Surface of the System, a (6 for the Cooke Lens)

$$
\begin{aligned}
H_a{}^U &= M_a = \partial U'/\partial U, \quad & K_a{}^U &= R_a = \partial L'/\partial U, \\
H_a{}^L &= O_a = \partial U'/\partial L, \quad & K_a{}^L &= T_a = \partial L'/\partial L, \qquad (13) \\
H_a{}^r &= P_a = \partial U'/\partial r, \quad & K_a{}^r &= V_a = \partial L'/\partial r.
\end{aligned}
$$

No coefficient is given for a change of index since the entire element must be treated as a unit. See next surface.

Next to the Last Surface of the System, $a-1$

$$
\begin{aligned}
H^U{}_{a-1} &= M_{a-1} M_a + R_{a-1} O_a, \\
H^L{}_{a-1} &= O_{a-1} M_a + T_{a-1} O_a, \\
H^r{}_{a-1} &= P_{a-1} M_a + V_{a-1} O_a, \\
H^{N/N'}{}_{a-1} &= Q_{12}, \text{ (See Table I)} \\
K^U{}_{a-1} &= M_{a-1} R_a + R_{a-1} T_a, \\
K^L{}_{a-1} &= O_{a-1} R_a + T_{a-1} T_a, \\
K^r{}_{a-1} &= P_{a-1} R_a + V_{a-1} T_a, \\
K^{N/N'}{}_{a-1} &= W_{12}. \text{ (See Table I)}
\end{aligned} \qquad (14)
$$

Third Surface From the Rear of the System, $a-2$

$$
\begin{aligned}
H^U{}_{a-2} &= M_{a-2} H^U{}_{a-1} + R_{a-2} H^L{}_{a-1}, \\
H^L{}_{a-2} &= O_{a-2} H^U{}_{a-1} + T_{a-2} H^L{}_{a-1}, \\
H^r{}_{a-2} &= P_{a-2} H^U{}_{a-1} + V_{a-2} H^L{}_{a-1}, \\
K^U{}_{a-2} &= M_{a-2} K^U{}_{a-1} + R_{a-2} K^L{}_{a-1}, \\
K^L{}_{a-2} &= O_{a-2} K^U{}_{a-1} + T_{a-2} K^L{}_{a-1}, \\
K^r{}_{a-2} &= P_{a-2} K^U{}_{a-1} + V_{a-2} K^L{}_{a-1}.
\end{aligned} \qquad (15)
$$

The element is treated as a whole as to index.

Fourth Surface From the Rear of the System, $a-3$

$$
\begin{aligned}
H^U{}_{a-3} &= h^U{}_{a-3} H^U{}_{a-1} + k^U{}_{a-3} H^L{}_{a-1}, \\
H^L{}_{a-3} &= h^L{}_{a-3} H^U{}_{a-1} + k^L{}_{a-3} H^L{}_{a-1}, \\
H^r{}_{a-3} &= h^r{}_{a-3} H^U{}_{a-1} + k^r{}_{a-3} H^L{}_{a-1}, \\
H^{N/N'}{}_{a-3} &= h^{N/N'}{}_{a-3} H^U{}_{a-1} + k^{N/N'}{}_{a-3} H^L{}_{a-1}, \\
K^U{}_{a-3} &= h^U{}_{a-3} K^U{}_{a-1} + k^U{}_{a-3} K^L{}_{a-1}, \\
K^L{}_{a-3} &= h^U{}_{a-3} K^U{}_{a-1} + k^L{}_{a-3} K^L{}_{a-1}, \\
K^r{}_{a-3} &= h^r{}_{a-3} K^U{}_{a-1} + k^r{}_{a-3} K^L{}_{a-1}, \\
K^{N/N'}{}_{a-3} &= h^{N/N'}{}_{a-3} K^U{}_{a-1} + k^{N/N'}{}_{a-3} K^L{}_{a-1}.
\end{aligned} \qquad (16)
$$

FIFTH SURFACE FROM THE REAR OF THE SYSTEM, $a-4$

$$
\begin{aligned}
H^U{}_{a-4} &= M_{a-4}H^U{}_{a-3} + R_{a-4}H^L{}_{a-3},\\
H^L{}_{a-4} &= O_{a-4}H^U{}_{a-3} + T_{a-4}H^L{}_{a-3},\\
H^r{}_{a-4} &= P_{a-4}H^U{}_{a-3} + V_{a-4}H^L{}_{a-3},\\
K^U{}_{a-4} &= M_{a-4}K^U{}_{a-3} + R_{a-4}K^L{}_{a-3},\\
K^L{}_{a-4} &= O_{a-4}K^U{}_{a-3} + T_{a-4}K^L{}_{a-3},\\
K^r{}_{a-4} &= P_{a-4}K^U{}_{a-3} + V_{a-4}K^L{}_{a-3}.
\end{aligned}
\tag{17}
$$

The element is treated as a whole as to index.

SIXTH SURFACE FROM THE REAR OF THE SYSTEM, $a-5$

$$
\begin{aligned}
H^U{}_{a-5} &= h^U{}_{a-5}H^U{}_{a-3} + k^U{}_{a-5}H^L{}_{a-3},\\
H^L{}_{a-5} &= h^L{}_{a-5}H^U{}_{a-3} + k^L{}_{a-5}H^L{}_{a-3},\\
H^r{}_{a-5} &= h^r{}_{a-5}H^U{}_{a-3} + k^r{}_{a-5}H^L{}_{a-3},\\
H^{N/N'}{}_{a-5} &= h^{N/N'}{}_{a-5}H^U{}_{a-3} + k^{N/N'}{}_{a-5}H^L{}_{a-3},\\
K^U{}_{a-5} &= h^U{}_{a-5}K^U{}_{a-3} + k^U{}_{a-5}K^L{}_{a-3},\\
K^L{}_{a-5} &= h^L{}_{a-5}K^U{}_{a-3} + k^L{}_{a-5}K^L{}_{a-3},\\
K^r{}_{a-5} &= h^r{}_{a-5}K^U{}_{a-3} + k^L{}_{a-5}K^L{}_{a-3},\\
K^{N/N'}{}_{a-5} &= h^{N/N'}{}_{a-5}K^U{}_{a-3} + k^{N/N'}{}_{a-5}K^L{}_{a-3}.
\end{aligned}
\tag{18}
$$

The law of progression is now evident, so that the equations can be applied to any number of surfaces.

One valuable point should be mentioned, namely, that it can be proven that the $a-2$ set of equations (third to the last surface) is far more general than it appears here. If changes in parameters occur only at one surface, then those equations can be used to get from one surface to the next, anywhere in the system, the $a-2$ subscripts referring to the surface in which the changes are taking place, and the $a-1$ subscript to the previous surface, counting from the rear. If, however, a change in any parameter in any surface requires a change in the other surface of any element, then the form $a-3$ must be used. For example, when a lens element is merely bent and thus both radii changed, or in every case where the glass is changed, the $a-3$ group must be used. This is applicable anywhere in the system. One need only remember what the letters stand for. This is sufficiently important to warrant repetition.

The capital H's are the coefficients *in the final image space* of the parameter changes for the terms making up the total change in the final vergence angles, and the capital K's are the corresponding coefficients of the terms making up the change in the final intersection lengths. The parameters to which these coefficients apply are indicated by the superscripts, and the surface where the changes are taking place by the subscripts. The lower case h's and k's are merely values obtained by applying the $a-1$ group of equations to that entire lens element, no matter where it happens to be in the optical system. If they are for the last element of the system, they automatically become capitals.

It will save much time in applying this method to any particular problem if the form of these equations be first carefully studied. There are four equations (except in the case of changes in the index of refraction. This equation appears only for even surfaces, since the entire element must be taken as a unit whenever the glass is changed). There is an equation for each of the parameters, U, L, r, and N/N'. In each equation there are always two terms, always added together and always made up in the same way.

(a) The first factor in each term comes from: the current surface if the surface is an odd one, counted from the back of the system; the current element if the surface is an even one. Note further that the first of these are vergence angle coefficients and the second intersecting length coefficients, and in each case these coefficients are those associated with the parameter under consideration.

(b) The second factor in each term is a *final image space* coefficient, already determined for the system up to the point under consideration. For the *new vergence angle coefficients* these second factors are also *vergence angle coefficients*, the first, the U coefficient and the same for all four equations, the second the L coefficient and the same for all four equations (one equation for each parameter, U, L, r, and N/N'). For the new *intersecting length coefficient*, these second factors are the *length coefficients* of the previous surface

calculations, a U coefficient in the first term and an L coefficient in the second, and again the same in all four equations.

Recognizing the above facts will greatly reduce the labor of computing. As a matter of fact, the process is exactly the same at all points in the system no matter how complex the system, so that, if this process is memorized, the equations are no longer necessary.

If a change in the amount of separation of two elements of a system are under consideration, the effect to be expected in the final image space is determined by simply assuming a change in the incident intersecting length (L) for the *first* surface following the expansion. dU, dr, $d(N/N')$ are then zero. If the thickness of the lens is to be changed, one simply assumes a change in the intersecting length for the *second* surface of that element. In either case, the adjustment coefficient for any particular aberration is obtained by substituting the $H_x{}^{L''}s$ and the $K_x{}^{L''}s$ (the final image space coefficients) into the expression for that aberration to be given later, where x is the surface where the dL occurs.

If both radii are to be changed in such a manner as to leave the focal length and the thickness of the element unchanged, that is, the element is merely "bent," then the extra term in Eqs. (9) and (10) fully covers the case, but, if only one radius of an element is changed, this extra term disappears, but extra terms will appear elsewhere. For instance, in the equation for the distortion, the focal length of the system appears. If, therefore, only one radius is changed, the focal length of the element will change according to the equation

$$1/f=(N-1)\{(1/r_1)-(1/r_2)+[(N-1)t/Nr_1r_2]\}. \quad (19)$$

N is the index of refraction, r_1 and r_2 the first and second radii of the element, and t the thickness of the element. This change in f will change the focal length of the system (F'').

After this new focal length of the element has been determined, the new focal length of the entire system can be calculated by using the proper equation for the system under consideration. For the Cooke lens examined in this article, (any system made up of three elements separated by air), the proper equation takes the following form,

$$\frac{1}{F''}=\frac{1}{f_1}+\frac{1}{f_2}+\frac{1}{f_3}-\frac{S_1(f_3+f_2)+S_2(f_1+f_2)-S_1S_2}{f_1f_2f_3}, \quad (20)$$

where f_1, f_2, and f_3 are the focal lengths of the three elements taken in order in the direction of light propagation, S_1 is the distance between the principal planes of elements one and two (posterior of element one and the anterior of element two), and S_2 the corresponding distance between elements two and three. Both S_1 and S_2 are always positive quantities.

The new method of correcting lens systems is now complete. The next step will be the determination of the equations giving the differential changes of the various aberrations in the final image space in terms of the changes in the parameters of the system.

GENERAL DIFFERENTIAL EQUATIONS

In the following equations, the lower case letters indicate the paraxial region and the upper case letters refer to rays in which the angles made with the axis of the system are too large to permit replacing the sines by the angles and cosines by one.

Conrady gives the following equations relating the various quantities in the final image space.

The height of the image formed by the top and bottom tangential rays above the axis of the optical system is

$$H_{ul}''=\frac{(L_l''-L_u'')\sin U_u\sin U_u''}{(\sin(U_u''-U_l''))}. \quad (21)$$

The distance, measured from the pole along the axis, of the intersection of the two tangential rays for a given aperture is

$$L_{ul}''=L_l''-\frac{(L_l''-L_u'')\sin U_u''\cos U_l''}{\sin(U_u''-U_l'')}, \quad (22)$$

where u, l, and, in later equations pr, indicate the upper, lower and the principal rays of a given aperture, respectively.

With the aid of these equations, all the aberration equations will now be collected, expressed, however, in terms of the end results of ray tracings.

$$\text{Long. Chro. Ab.} = l_y'' - l_v'', \tag{23}$$

where y refers to the longer wave-length, v to the shorter. This quantity is positive when blue (v) is nearest to the lens.

$$\text{Trans. Chro. Ab.} = (L_{y_{pr}}'' - L'') \tan U_{y_{pr}}'' - (L_{v_{pr}}'' - L'') \tan U_{v_{pr}}'' \tag{24}$$

where L'' is the axial intersection of a marginal ray, and $L_{y_{pr}}''$ and $U_{y_{pr}}''$ refer to the y component of the principal oblique ray and $L_{v_{pr}}''$ and $U_{v_{pr}}''$ to the v component of the principal oblique ray. This aberration is positive when red is away from the axis.

$$\text{Sph. Ab.} = l'' - L''. \tag{25}$$

This is positive when the marginal ray intersection is nearest the lens.

$$\text{Coma}_T' = (L_{pr}'' - L_{ul}'') \tan U_{pr}'' - H_{ul}''$$

$$= \left(L_{pr}'' - L_l'' + \frac{(L_l'' - L_u'') \sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')} \right) \tan U_{pr}'' - \frac{(L_l'' - L_u'') \sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}. \tag{26}$$

This is positive when the coma points away from the axis.

$$\text{Curvature} = (L_{ul}'' - L'') = L_l'' - L'' - \frac{(L_l'' - L_u'') \sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')}. \tag{27}$$

This is positive when concave in the direction of light propagation.

$$\text{Distortion} =$$

for near objects

$$= H_0 \frac{N_0 u_0}{N'' u''} - (L_{pr}'' - l'') \tan U_{pr}'', \tag{28}$$

for distant objects

$$= -\frac{N_0}{N''} f'' \tan U_{pr_0} - (L_{pr}'' - l'') \tan U_{pr}''. \tag{29}$$

This is positive if the ideal image is larger than the actual image.

Astigmatism cannot be determined definitely without tracing sagittal rays, but if the obliquity is not excessive and if there are no very deeply curved surfaces in the system, an estimate can be made by taking $\frac{3}{4}$ the difference between the tangential curvature and the Petzval curvature (see Conrady p. 288).

Petzval curvature

$$= \tfrac{1}{2} N_k H_k^2 \sum \frac{N_j' - N_j}{N_j' N_j r} \text{ surface by surface} \tag{30}$$

$$= \tfrac{1}{2} H_k^2 \sum \frac{N_j - 1}{N_j} C_j \text{ lens by lens.} \tag{31}$$

No equations were developed for astigmatism in this article because the above approximation is usually sufficient and is easily derived from the above fundamental equation and the curvature equations already determined. In general the Petzval curve need be calculated only once since it is not affected by either bending the lenses or by changing their thicknesses or separation. The *change* in the astigmatism is then near enough equal to $\frac{3}{4}$ the change in tangential curvature.

Differentiating each of the above equations and collecting terms involving dU_{pr}'', dU_u'', dU_l'', dL_{pr}'', dL_u'', dL_l'', we arrive at the following,

---

$$d(\text{Long. Chro. Ab.}) = dl_y'' - dl_v'',$$

$$d(\text{Trans. Chro. Ab.}) = (L_{y_{pr}}'' - L'') \sec^2 U_{y_{pr}}'' dU_{y_{pr}}'' - (L_{v_{pr}}'' - L'') \sec^2 U_{v_{pr}}'' dU_{v_{pr}}'' + \tan U_{y_{pr}}'' (dL_{y_{pr}}'' - dL'') - \tan U_{v_{pr}}'' (dL_{v_{pr}}'' - dL''),$$

$$d(\text{Sph. Ab.}) = dl'' - dL'',$$

$$d(\text{Coma}_T') = \left[L_{pr}'' - L_l'' + \frac{(L_l'' - L_u'')\sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')}\right]\sec^2 U_{pr}'' dU_{pr}''$$

$$+\left[\tan U_{pr}'' \frac{(L_l'' - L_u'')\cos U_u'' \cos U_l''}{\sin(U_u'' - U_l'')} - \tan U_{pr}''\right.$$

$$\times \frac{(L_l'' - L_u'')\sin U_u'' \cos U_l'' \cot(U_u'' - U_l'')}{\sin(U_u'' - U_l'')} - \frac{(L_l'' - L_u'')\cos U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}$$

$$\left.+\frac{(L_l'' - L_u'')\sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}\cot(U_u'' - U_l'')\right] dU_u''$$

$$+\left[-\tan U_{pr}'' \frac{(L_l'' - L_u'')\sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')} + \tan U_{pr}''\right.$$

$$\times \frac{(L_l'' - L_u'')\sin U_u'' \cos U_l'' \cot(U_u'' - U_l'')}{\sin(U_u'' - U_l'')} - \frac{(L_l'' - L_u'')\sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')}$$

$$\left.-\frac{(L_l'' - L_u'')\sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}\cot(U_u'' - U_l'')\right] dU_l'' + \tan U_{pr}'' dL_{pr}''$$

$$+\left[-\tan U_{pr}'' \frac{\sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')} + \frac{\sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}\right] dL_u''$$

$$+\left[+\tan U_{pr}'' \frac{\sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')} - \frac{\sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')} - \tan U_{pr}''\right] dL_l'',$$

$$d(\text{curvature}) = \left[\frac{(L_l'' - L_u'')\sin U_u'' \cos U_l'' \cot(U_u'' - U_l'')}{\sin(U_u'' - U_l'')}\right.$$

$$\left.-\frac{(L_l'' - L_u'')\cos U_u'' \cos U_l''}{\sin(U_u'' - U_l'')}\right] dU_u'' + \left[\frac{(L_l'' - L_u'')\sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}\right.$$

$$\left.-\frac{(L_l'' - L_u'')\sin U_u'' \cos U_l'' \cot(U_u'' - U_l'')}{\sin(U_u'' - U_l'')}\right] dU_l'' - dL'$$

$$+\left[\frac{\sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')}\right] dL_u'' + \left[1 - \frac{\sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')}\right] dL_l''.$$

Distortion

In the two equations for distortion, one for near objects and one for far objects, the N_0 and N'' refer to the first and last media, which in most cases are air and thus cancel out. In any case they are not involved in the differentiating. U_{pr0} is fixed by the desired angle of view of the system. F'' is the focal length of the entire system and is generally held constant, so that in the differentiation this term disappears. Since changes in F'' are sometimes made this term will be retained, but in this case F'' must be differentiated with respect to either r or L before the differential adjustment coefficients can

be used. With this understanding we will write

$$d(\text{distortion}) = -(L_{pr}'' - l_{par}'')\sec^2 U_{pr}'' dU_{pr}'' - \tan U_{pr}''(dL_{pr}'' - dl_{par}'') - H_0 \frac{N_0 u_{0par}}{N'' u_{par}''^2} du_{par}''$$

for near objects,

$$= -\frac{N_0}{N''}\tan U_{pr0}'' dF'' - (L_{pr}'' - l_{par}'')\sec^2 U_{pr}'' dU_{pr}'' - \tan U_{pr}''(dL_{pr}'' - dl_{par}'')$$

for distant objects.

CONDENSING THE ABOVE DIFFERENTIAL EQUATIONS

It will be desirable at this point to set down a group of equations whose greatest usefulness is in connection with the computing, but which are introduced at this point to simplify the form of the rather cumbersome equations resulting from the differentiation.

$$\begin{array}{cccc} L_1'' - L_u'' & \tan U_{pr}'' & \sin U_u'' & \cos U_u'' & \sin(U_u'' - U_l'') \\ & \sec^2 U_{pr}'' & \sin U_l'' & \cos U_l'' & \cot(U_u'' - U_l'') \end{array}$$

$$\alpha = \frac{\sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}, \quad A = \frac{(L_l'' - L_u'')\sin U_u'' \sin U_l''}{\sin(U_u'' - U_l'')},$$

$$\beta = \frac{\cos U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}, \quad B = \frac{(L_l'' - L_u'')\cos U_u'' \sin U_l''}{\sin(U_u'' - U_l'')}, \tag{32}$$

$$\gamma = \frac{\sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')}, \quad C = \frac{(L_l'' - L_u'')\sin U_u'' \cos U_l''}{\sin(U_u'' - U_l'')},$$

$$D = \frac{(L_l'' - L_u'')\cos U_u'' \cos U_l''}{\sin(U_u'' - U_l'')}.$$

The preceding set of equations may then be written

$$d(\text{Long. Chro. Ab.}) = dl_y'' - dl_v''. \tag{33}$$

$$d(\text{Trans. Chro. Ab.}) = (L_{y_{pr}}'' - L'')\sec^2 U_{y_{pr}}'' dU_{y_{pr}}'' - (L_{v_{pr}}'' - L'')\sec^2 U_{v_{pr}}'' dU_{v_{pr}}'' + \tan U_{y_{pr}}''(dL_{y_{pr}}'' - dL'') - \tan U_{v_{pr}}''(dL_{v_{pr}}'' - dL''). \tag{34}$$

$$d(\text{Sph. Ab.}) = dl'' - dL''. \tag{35}$$

$$\begin{aligned} d(\text{Coma}_T) = {} & [L_{pr}'' - L_l'' + C]\sec^2 U_{pr}'' dU_{pr}'' \\ & + [D\tan U_{pr}'' - C\tan U_{pr}''\cot(U_u'' - U_l'') - B + A\cot(U_u'' - U_l'')]dU_u'' \\ & + [-A\tan U_{pr}'' + C\tan U_{pr}''\cot(U_u'' - U_l'') - C - A\cot(U_u'' - U_l'')]dU_l'' \\ & + \tan U_{pr}'' dL_{pr}'' + [-\gamma\tan U_{pr}'' + \alpha]dL_u'' \\ & + [\gamma\tan U_{pr}'' - \alpha - \tan U_{pr}'']dL_l'', \text{ when } U_b'' > U_a'', \end{aligned} \tag{36}$$

or

$$\begin{aligned} = {} & [L_{pr}'' - L_u'' + B]\sec^2 U_{pr}'' dU_{pr}'' \\ & + [-A\tan U_{pr}'' - B\tan U_{pr}''\cot(U_u'' - U_l'') - B + A\cot(U_u'' - U_l'')]dU_u'' \\ & + [D\tan U_{pr}'' + B\tan U_{pr}''\cot(U_u'' - U_l'') - C - A\cot(U_u'' - U_l'')]dU_l'' \\ & + \tan U_{pr}'' dL_{pr}'' + [-\tan U_{pr}'' - \beta\tan U_{pr}'' + \alpha]dL_u'' \\ & + [\beta\tan U'' - \alpha]dL_l'', \text{ when } U_a'' > U_b''. \end{aligned} \tag{37}$$

Two forms have just been given for $d(\text{Coma}_T)$. They should give the same values but each is a little more accurate for the specified conditions.

$$d(\text{curvature}) = [C\cot(U_u'' - U_l'') - D]dU_u'' + [A - C\cot(U_u'' - U_l'')]dU_l'' - dL'' + \gamma dL_u'' + [1-\gamma]dL_l''. \quad (38)$$

$$d(\text{distortion}) = -(L'' - l'')\sec^2 U_{pr}''dU_{pr}'' - \tan U_{pr}''(dL_{pr}'' - dl'') - H_0\frac{Nu_{par}}{N'u_{par}'}du_{par}', \text{ for near objects,} \quad (39)$$

$$= -(N_0/N'')\tan U0_{pr}df' - (L_{pr}'' - l'')\sec^2 U_{pr}''dU_{pr}'' - \tan U_{pr}''(dL'' - dl''), \text{ for distant objects.} \quad (40)$$

Substituting into these equations the corresponding differentials from Eqs. (9) and (10) we get, for each of the variables in the element to be changed (here represented by x), the following final equations for determining the amount of the variations to be expected in each of the aberrations, brought about by the variations of the variables, dU_1, dL_1, (or dL_2 as previously discussed), dr_1, (or dr_2 as previously discussed), and dN'.

$$d(\text{Long. Chro. Ab.}) = (K_{ly}{}^x - K_{lv}{}^x)dx, \quad (41)$$

where K_{ly} and K_{lv} are the K coefficients of the new method for the paraxial region and for the longer and shorter wave-lengths respectively, and dx is the change in the parameter under examination, that is whose effect in the final image space is being sought. All the K's and H's and the double-primed letters are for the final image space. See notation earlier in this article.

$$d(\text{Trans. Chro. Ab.}) = \{(L''_{y_{pr}} - L'')\sec^2 U''_{y_{pr}}H^x{}_{y_{pr}} - (L''_{v_{pr}} - L'')\sec^2 U''_{pr}H^x{}_{v_{pr}} + \tan U''_{y_{pr}}(K^x{}_{y_{pr}} - K^x{}_{L''}) - \tan U''_{v_{pr}}(K^x{}_{v_{pr}} - K^x{}_{L''})\}dx. \quad (42)$$

$$d(\text{Sph. Ab.}) = (K^x{}_{l''} - K^x{}_{L''})dx. \quad (43)$$

$$d(\text{Coma}_T) = \{[L''_{pr} - L''_l + C]\sec^2 U''_{pr}H^x{}_{pr} + [D\tan U''_{pr} - C\tan U''_{pr}\cot(U''_u - U''_l) - B + A\cot(U''_u - U''_l)]H^x{}_u + [-A\tan U''_{pr} + C\tan U''_{pr}\cot(U''_u - U''_l) - C - A\cot(U''_u - U''_l)]H^x{}_l + \tan U''_{pr}K^x{}_{pr} + [-\gamma\tan U''_{pr} + \alpha]K^x{}_u + [\gamma\tan U''_{pr} - \alpha - \tan U''_{pr}]K^x{}_l\}dx. \quad (44)$$

$$d(\text{curvature}_T) = \{[C\cot(U''_u - U''_l) - D]H^x{}_u + [A - C\cot(U''_u - U''_l)]H^x{}_l - K^x{}_{L''} + \gamma K^x{}_u + [1-\gamma]K^x{}_l\}dx. \quad (45)$$

$$d(\text{distortion}) = \left\{-(L''_{pr} - l'')\sec^2 U''_{pr}H^x{}_u - \tan U''_{pr}(K^x{}_{pr} - K^x{}_l) - H\frac{Nu_{par}}{N''u''^2{}_{par}}H^x{}_u\right\}dx, \text{ for near objects,} \quad (46)$$

$$= -\frac{N_0}{N''}\tan U0_{pr}dF''{}^{\ddagger} - \{(L''_{pr} - l'')\sec^2 U''_{pr}H^x{}_{pr} + \tan U''_{pr}(K^x{}_{pr} - K^x{}_{l''})\}dx, \text{ for distant objects.} \quad (47)$$

‡ Note: This term can be evaluated in terms of one of the regular parameters, (r and N) through the use of the lens Eqs. (19) and (20). More often the focal length of the element is maintained and so this term drops out.

ERROR CALCULATIONS AS APPLIED TO RAY TRACINGS

It frequently becomes necessary to examine ray tracings for possible serious losses in precision.

The determining of the probable errors which creep into ray tracing data is not simple, since the errors which have to be combined are generally not independent. Furthermore the number of independent errors are not sufficiently numerous to warrant a complete analysis based on the laws of statistics and probabilities. Therefore, the author has chosen to offer here an error analysis, especially applicable to the ray tracing procedure. While it leaves much to be desired, it has reconciled the discrepancies which have here arisen.

The method starts with the standard ray tracing equations

$$U' = U + I - I', \tag{1e}$$

and

$$L' - r = (N/N')(\sin U/\sin U')(L - r). \tag{2e}$$

From the second equation it will be clear that, while $L'-r$ will depend on $\sin U'$, $\sin U'$ itself will depend on N/N', $\sin U$, and $(L-r)$. The interdependence of these quantities makes the calculating of errors in this case quite involved, and not the simple matter which independent quantities and Eq. (2e) above might suggest.

Suppose the second equation is differentiated to obtain

$$\begin{aligned} d(L'-r) &= \frac{\partial(L'-r)}{\partial N/N'} d\frac{N}{N'} + \frac{\partial(L'-r)}{\partial \sin U} d\sin U \\ &\quad + \frac{\partial(L'-r)}{\partial(L-r)} d(L-r) - \frac{\partial(L'-r)}{\partial \sin U'} d\sin U' \\ &= \frac{(L-r)\sin U}{\sin U'} d\frac{N}{N'} + \frac{N}{N'}\frac{(L-r)}{\sin U'} d\sin U \\ &\quad + \frac{N}{N'}\frac{\sin U}{\sin U'} d(L-r) - \frac{N}{N'}\frac{(L-r)}{\sin^2 U'} d\sin U'. \end{aligned}$$

If this equation is divided by $(L'-r)$ to change the terms to "fractional parts" or "parts in one" (not percentage, the former is more convenient), it becomes

$$\begin{aligned} d(L'-r)/L'-r &= d\frac{N}{N'} \Big/ \frac{N}{N'} + d\sin U/\sin U \\ &\quad + d(L-r)/L-r - d\sin U'/\sin U', \end{aligned}$$

and if we now look at the differential quantities as errors, this equation may be written, where 1_1 stands for "parts in one,"

$$\begin{aligned} 1_1E_{L'-r} &= 1_1E_{N/N'} + 1_1E_{\sin U} \\ &\quad + 1_1E_{L-r} - 1_1E_{\sin U'}. \end{aligned} \tag{3e}$$

Note that the negative sign on the right must be retained since the error in this term must be of opposite sign from the other three because the latter are in the numerator of (2e) while $1_1E_{\sin U'}$ is in the denominator and dependent on the three in the numerator. As a matter of fact, a single sign, and not $\pm$ had better be used (one could cover the above condition by writing $\pm$ and $\mp$) since the errors sometimes change sign in going from one surface to the next or from one stage of the ray tracing to another. This condition also necessitates careful study each time a new error enters the calculation to make sure that the effect of the new error is in the correct direction when combined with the errors already there. The signs of the quantities are completely ignored as they are fully accounted for by the above treatment whenever they enter the problem.

The use of Eq. (3e) is not simple. If we refer to Form No. 1 we note the following facts:

1. The error in r enters only by way of the term $(L-r)$. In the ray tracing calculation the error in r appearing in item 7 is completely eliminated by the same error in item 11.

2. It would seem (referring to items 16 and 18) that the errors in these items cancel. This is not true. This step is one of the most annoying sources of the loss of precision in ray tracing calculations. The errors will cancel of course, if the angles are equal, which is rarely the case. Any errors appearing here will finally show up in item 19 or U'.

3. Since N/N' appears only once (item 9), the error in N/N' should be introduced here, thus added to the errors in I to give the error in I'. The term $1_1E_{N/N'}$ in Eq. (3e) is then omitted, as it means little where it is and is fully accounted for when introduced at item 9.

4. If we examine the check equation

$$L' = \frac{L\sin U \cos 1/2(I'-U')}{\sin U' \cos 1/2(I-U)}, \tag{4e}$$

we find that after U' and its error have been de-

termined, the error equation

$$1_1E_{L'-r}=1_1E_{\sin U}+1_1E_{L-r}-1_1E_{\sin U'} \quad (5e)$$

used in the original calculation is applicable to the check (4e), if we substitute 1_1E_L for 1_1E_{L-r}. The errors in the cosines are small and may in general be neglected.

Before Eq. (5e) can be solved, U' and $1_1E_{\sin U'}$ must be determined. This can be worked out directly from the ray tracing data, but perhaps is more easily calculated from an equation derived as follows;

differentiating

$$U'=U+I-I'$$

gives

$$dU'=dU+dI-dI',$$

or

$$\frac{dU'}{U'}=\frac{U}{U'}\frac{dU}{U}+\frac{I}{U'}\frac{dI}{I}-\frac{I'}{U'}\frac{dI'}{I'},$$

and, using the notation previously used for "parts in one,"

$$1_1E_{U'}=\frac{U}{U'}1_1E_U+\frac{I}{U'}1_1E_I-\frac{I'}{U'}1_1E_{I'}. \quad (6e)$$

This equation will be more useful if expressed in terms of the sines and log sines of the angles.

Differentiating a sine gives,

$$d\sin x=\cos x dx,$$
$$1_1E_{\sin x}=x\cot x 1_1E_x.$$

Differentiating a log sine gives,

$$d\log\sin x=(\log_a e/\sin x)d\sin x,$$

$$1_1E_{\log\sin x}=\frac{\log_a e}{\log\sin x}1_1E_{\sin x}\ (a \text{ is usually } 10).$$

Substituting back into (6e) gives

$$\begin{aligned}1_1E_{\sin U'}&=\frac{U}{\tan U'}1_1E_U+\frac{I}{\tan U'}1_1E_I-\frac{I'}{\tan U'}1_1E_{I'}\\&=\frac{\tan U}{\tan U'}1_1E_{\sin U}+\frac{\tan I}{\tan U'}1_1E_{\sin I}-\frac{\tan I'}{\tan U'}1_1E_{\sin I'}\\&=\frac{\log\sin U\tan U}{\log_{10}e\tan U'}1_1E_{\log\sin U}+\frac{\log\sin I}{\log_{10}e}\frac{\tan I}{\tan U'}1_1E_{\log\sin I}-\frac{\log\sin I'}{\log_{10}e}\frac{\tan I'}{\tan U'}1_1E_{\log\sin I'}. \quad (7e)\end{aligned}$$

Any of the corresponding terms in any of the right-hand members may be used at any time depending upon the data available.

Or we may find the following form even more useful,

$$\begin{aligned}1_1E_{\log\sin U'}&=\frac{\log\sin U}{\log\sin U'}1_1E_{\log\sin U}+\frac{\log\sin I}{\log\sin U'}\frac{\tan I}{\tan U'}1_1E_{\log\sin I}-\frac{\log\sin I'}{\log\sin U'}\frac{\tan I'}{\tan U'}1_1E_{\log\sin I'}\\&=\frac{E_{\log\sin U}}{\log\sin U'}+\frac{E_{\log\sin I}}{\log\sin U'}\frac{\tan I}{\tan U'}-\frac{E_{\log\sin I'}}{\log\sin U'}\frac{\tan I'}{\tan U'}. \quad (8e)\end{aligned}$$

THE PROCEDURE IN ERROR CALCULATIONS

The errors in U' and L' are then calculated as follows:

The absolute error in item 1 (Form 1) is added to the absolute error in item 2 to get the error in item 3.

One more equation is needed in getting from item 3 to item 5:

$$d\log x=(dx/x)\log_a e.$$

Absolute errors in items 4 and 5 are added algebraically to get the absolute error in item 6 which will be transferred unchanged to 8 and 10 unless there is an error in N/N' which must then be introduced here with the proper regard for signs. Errors in r are ignored as previously explained.

It is now only necessary to make use of Eqs. (8e) then (5e) and transfer these values to the next surface and repeat the steps.

In making these error calculations on a set of ray tracing data, it seems, in general, to be satisfactory to assume that the half-point doubt in the last place may be ignored, on the basis that, for the number of separate items used in the calculations, they may be assumed to cancel to such a degree as to leave small errors compared to the accumulated errors sought. However, at the first surface these errors are assumed for starting errors, and when the above procedure results in an error less than the half point in the last place, the half-point error is restored. The above procedure accounts for only error accumulations, those which grow by virtue of the ray tracing procedure, and do not take into account the new errors continually introduced by rounding off of final digits, or the introduction of new terms.

FORM NO. I.

Project No. ..... Test No. ..... Surface No. .....

Item	Quantity	Marginal ray $N=$	Paraxial ray $N=$
1	L		
2	R		
3	$L-R$		
4	$+\log \sin U$		
5	$+\log(L-R)$		
6	$\log(L-R) \sin U$		
7	$-\log R$		
8	$\log \sin I$		
9	$+\log N/N'$		
10	$\log \sin I'$		
11	$+\log R$		
12	$\log R \sin I'$		
13	$-\log \sin U'$		
14	$\log(L'-R)$		
15	U		
16	I		
17	$U+I$		
18	$-I'$		
19	U'		
20	$\frac{1}{2}(I+U)$		
21	$\frac{1}{2}(I-U)$		
22	$\frac{1}{2}(I'-U')$		
23	$L'-R$		
24	$+R$		
25	L'		
26	$+\log L$		
27	$+\log \sin U$		
28	$\log L \sin U$		
29	$-\log \cos \frac{1}{2}(I-U)$		
30	$\log PA$		
31	$-\log \sin U'$		
32	$\log PA/\sin U'$		
33	$+\log \cos \frac{1}{2}(I'-U')$		
34	$\log L'$		
35	L'		
36	D'		
37	$L+$		
38	PA		
39	$X=\frac{1}{2}PA^2/R$		
40	$Y=PA \cos \frac{1}{2}(I+U)$		

ADVANTAGES OF THIS METHOD

Time

The total amount of time required to finish a complete design is in general considerably less than that required for the usual successive-ray-tracing method.

Effects Instantly Visible

The effect which each change in any of the numerous parameters will have on each of the various aberrations is instantly visible, so that one always knows if an anticipated change, designed to correct some aberration, is going to change some other aberration by an unacceptable amount.

All Corrections Made in One Operation

The table which is compiled by this method gives the effect which a change in each parameter has on each aberration, so that a quick glance at this table and some rapid and simple calculations will determine the best final compromise for any particular purpose without the mass of ray tracings and resulting computations.

Accuracy

The results obtained by this method are for more trustworthy than those obtained from ray tracings for the following reason. This method uses the left-hand digits (44.75635) of the ray tracing data while the aberrations are obtained from the right-hand digits (44.75635). As ray tracing proceeds the lack of sufficient precision moves the doubt in as much as three and even more significant figures from the last right hand digit, while the left hand digits will not be effected.

No approximations have been made in deriving the coefficients presented by this method. However, objections have been raised to the effect that the functions differentiated in arriving at these results are not straight line functions. This was fully realized from the start, but it was also realized that the functions are, at least to the first order, wide parabolas. It was therefore believed that the amount of discrepancy introduced would not be serious for the amounts the parameters are likely to be changed in making the final adjust-

ments of an optical system. This has turned out to be the case as will be seen when the test data is presented. It should be noted that, since a differential coefficient at a point of a curve approaches that of a straight line through that point as the differential change approaches zero, the method here presented improves as the refinement advances, while algebraic approaches get worse for the same conditions, and sometimes become quite useless before the full refinement has been reached.

Tolerances

The differential adjustment coefficients give directly the tolerances for each parameter and thus make it possible to decide without additional calculations whether the lens system is practical, that is, capable of being manufactured.

APPLICATION OF THE NEW METHOD OF ANALYSIS TO A COOKE ANASTIGMAT, SERIES IV

Before offering an example of how this new method is applied and how well it works out in practice, attention should be called to the fact that the efficiency with which the work proceeds will depend upon the detailed procedure which is followed. The procedure will, therefore, be given in considerable detail together with all the tabulating and calculating forms. The steps will also be more easily followed with these forms as references.

Step One

The method here presented is for refining a design already past the preliminary stages. A preliminary design, roughly determined with the aid of first-order equations will be assumed. It is further assumed that a set of rays for determining the aberrations has been traced through the system. Form No. I is a very useful form for this purpose. Of course the columns have to be extended to accommodate the rest of the rays required. This form is for log computation and, while useful to those who have no computing machine at their command, is given this way here because it will be useful in the discussions. If the work is done with the aid of a calculating machine, this form can be much reduced. The only items in Form No. 1 which will then be needed are: 1, 2, 3, 4, 8, 9, 10, 13, 15, 16, 17, 18, 19, 23, 24, 25, 36, and 37. If the check is to be made and x and y are to be calculated, then items 20, 21, 22, 29, 33, 35, 39, and 40 will have to be retained. Of course the quantities, not the logs, are used with the calculating machine.

FORM NO. II.

Project No. ..... Test No. ..... Surface No. .....

L	$E=r\sin I'/\sin U'$
r	$F=(N/N')\sin U/\sin U'$
$L-r$	$G=\sin(I-I')/\cos I\cos I'$
N/N'	$E\cot U'$
U	
U'	
I	
I'	$M=L+G\cot U$
$I-I'$	$O=G/(L-r)$
$\sin(I-I')$	$P=-LG/r(L-r)$
$\sin U$	$Q=-\sin I/\cos I'$
$\sin U'$	
$\sin I$	
$\sin I'$	
$\cot U$	
$\cot U'$	$R=E\cot U-EM\cot U'$
$\cos I$	$T=F-EO\cot U'$
$\cos I'$	$V=1-F-EP\cot U'$
$\cos I\cos I'$	$W=EN'/N-EQ\cot U'$

Note: It should be kept in mind while computing this sheet that: (1) The expression here given for Q or W is correct only when changes in parameters are being made at the first refracting surface of an element. See Table I. (2) None of the above expressions are correct for plane refracting surfaces or for paraxial rays. See Table I.

Step Two

Items 1 to 8 inclusive and 11 to 14 inclusive, on Form No. II are now filled in from Form No. I. The remaining items on Form No. II are computed from those values just entered. In some cases a slide rule may be sufficient for these calculations, but in more complex systems three, sometimes four, and in rare cases five places are lost as a result of the subtraction of almost equal quantities. Often these cases are not too serious since the aberrations or the changes in the aberrations are so small that the serious doubt is of little consequence. Form No. II is made out for each surface and each ray involved in the calculation. The last eight items in this form are required in the further procedure. While the values can be taken from these forms, it will, in the end, save time and mistakes if the eight items, M, O, P, Q, R, T, V, and W are transferred to a new form, Form No. III.

The values required to fill out Form No. IV are calculated with the aid of the values on Form No. III and Eqs. (13) to (18). Cemented doublets are considered to be two lenses with zero separation.

FORM No. III.

Project No. ..... Test No. ..... Ray .....

Item	Surface 1	Surface 2	Surface 3	Surface 4	Surface 5	Surface 6
M						
O						
P						
Q						
R						
T						
V						
W						

FORM No. IV.

Project No. ..... Test No. ..... Ray .....

Item	Surface 1	Surface 2	Surface 3	Surface 4	Surface 5	Surface 6
H^U						
H^L						
H^r						
H^N						
K^U						
K^L						
K^r						
K^N						

From Eq. (13) it will be observed that the sixth column will be identical in Form III and Form IV and thus the values in Form No. III can be transferred to Form No. IV.

The quantities to be entered into Form No. IV are the rates of change of U and L in the *final image space* with changes in *each of the parameters of the system*. For instance, $H^{r}{}_{a-3}$ represents $\partial U''/\partial r_{a-3}$ that is, the change of divergence angle after refraction at the last surface with a change in the radius of the refracting surface fourth in from the rear end of the system, and $K^{r}{}_{a-3}$ represents $\partial L''/\partial r_{a-3}$ or the change in the intersecting length in the *final image space* with a change in r_{a-3}.

Substituting the quantities in Form No. IV into Eqs. (41) to (47) gives the quantities for Form No. V.

These substitutions offer no particular difficulties, but since the expression for distortion contains a term which requires special treatment, this case will be offered as an illustration.

SAMPLE CALCULATION
(Distortion)

The equation for distortion (47), adapted to a change in r_3 is

$$d(\text{dist.}) = -\frac{N_0}{N''}\tan U_{0pr}\,dF'' - \{(L_{pr}'' - l'')\sec^2 U_{pr}'' + \tan U_{pr}''(K_{pr}{}^{r_3} - K_{pr}{}^{r_3})\}dr_3.$$

The following values are obtained from the original tracing data.

$U_{0pr} = +24°\ 45'\ 00''$ ($\frac{1}{2}\times 0.707$ $\times$ the designer's claim of a field of 70°),
$U_{pr}'' = +25°\ 38'\ 36.85''$,
$L_{pr}'' = -12.48953$,
$l'' = +84.76539$.

From these we get

$$\tan U_{0pr} = +0.4610063,$$
$$\tan U_{pr}'' = +0.4800550,$$
$$\sec^2 U_{pr}'' = +1.230453.$$

From Table IV

$$H_{pr}{}^{r_3} = -0.00047664,$$
$$K_{pr}{}^{r_3} = +0.013898,$$
$$K_{l''}{}^{r_3} = +1.30567.$$

Before substituting in the d(dist.) equation it will be necessary to change dF'' (the term which requires special treatment) to dr_3.

Since dr_3 is the first surface of the second element of the system, Eq. (19) must be differentiated with respect to r_3,

$$\frac{1}{f_2} = (N-1)\left\{\frac{1}{r_3} - \frac{1}{r_4} + \frac{(N-1)t}{Nr_3r_4}\right\},$$

$$\frac{\partial f_2'}{\partial r_3} = \frac{f_2'^2}{r_3^2}\left\{(N-1)\left[1 + \frac{(N-1)t}{Nr_4}\right]\right\}.$$

Since this change in the focal length of f_2 changes the focal length of the system, it becomes necessary to differentiate Eq. (20).

$$\frac{1}{F''} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{1}{f_3} - \left\{\frac{S_1(f_3+f_2) + S_2(f_1+f_2) - S_1S_2}{f_1f_2f_3}\right.$$

with respect to f_2 which gives

$$\frac{\partial F''}{\partial f_2}=\frac{F''}{f_2{}^2}\left\{1+\frac{S_1S_2-S_1f_3-S_2f_1}{f_1f_3}\right\}.$$

Combining these gives

$$\frac{\partial F''}{\partial r_3}=\frac{F''}{r_3{}^2}\left\{1+\frac{S_1S_2-S_1f_3-S_2f_1}{f_1f_3}\right\}\times\left\{(N-1)\left[1+\frac{(N-1)t}{Nr_4}\right]\right\}.$$

The focal length of the system is

$$F''=\frac{\text{chosen height at first surface}}{\sin U''\text{ for the image space}}=\frac{9}{5^\circ\,4'\,36.76''}=101.3048,$$

$$\frac{\partial F''}{\partial r_3}=\text{(when indicated substitutions are made)}=0.01251840.$$

We are now ready to substitute in the d(dist.) equation. Recalling that object and image are in air and therefore $N_0=N''=1$

$$d(\text{distortion})=\{-0.4610063\times0.01251840-0.05703884+0.6201221\}dr_3=+0.5573122dr_3.$$

The full seven places are given in the above example to show how far the work was carried out. It will be noted later that in many cases these seven places were reduced to as low as three through losses in precision. In general not more than two or three places are required. It will be noted that the change in focal length of the system accounts for only 10 percent of the adjustment coefficient and so the above corrections can be ignored in many cases.

The above is just one of the coefficients of table Form No. V. By further substitutions in Eqs. (41) to (47) the rest of the coefficients are obtained.

The use of this table, Form No. V, is then as follows. One goes down the column of that aberration most in need of correction until the largest coefficient is found. The change in this parameter required to bring about the desired result is determined. The changes in all aberrations resulting from this change in parameter are computed and tabulated. If, on inspection of the results just obtained, any other aberration has become undesirably large, we go back to Form No. V and pass up and down the vertical column for this aberration to find a large coefficient with which this aberration can be brought back to an acceptable value without cancelling the correction already made. Again calculate the changes in all aberrations and tabulate, preferably under the pre-

FORM NO. V. Adjustment coefficients for the Cooke anastigmat series IV.

Parameter	Longitudonal Chro. Ab.	Transverse Chro. Ab.	Spherical aberration	Coma	Longitudinal curvature	Astigmatism	Distortion
L_1	0.00	+0.0080	0.0				+0.012
r_1	−0.50	−0.20	−10.0				+7.00
$\left(\frac{N}{N'}\right)_1$							
L_2	−0.10	−0.022	−3.5				+2.00
r_2	+0.003	+0.0005	+0.080				−0.15
L_3	−1.00	−0.25	−4.0				+5.0
r_3	−0.027030	−0.002668	−0.617878	−0.0854662	+1.11638		+0.5573122
$\left(\frac{N}{N'}\right)_3$							
L_4	−0.05	−0.0080	−1.5				+1.0
r_4	+0.30	+0.040	+5.0				+6.0
L_5	+0.0010	+0.050	+0.0040				+0.060
r_5	−0.0004	+0.000050	−0.0030				+0.010
$\left(\frac{N}{N'}\right)_5$							
L_6	+0.005	+0.0090	+0.010				+0.0090
r_6	+0.030	−0.0010	+0.090				−0.60

The values that apply to the Cooke anastigmat are only those given for r_3. All others are of a general order of magnitude only, and are inserted for illustrative purposes only.

TABLE II. Summary of results.

	Long. chr.	Tran. chr.	Sph. ab.	Coma	Curvature	Distortion
Parameter	r_3	r_3	r_3	r_3	r_3	r_3
Adj. coef.	−0.027030	−0.002668	−0.617878	−0.0854662	+1.11638	+0.5573122
Original aberration	+0.18034	−0.02893	−0.20773	+0.57210	−5.11347	−0.01445
After a 2% change in r_3	+0.21010	−0.02600	+0.46479	+0.66904	−6.30255	−0.63740
After a 20% change in r_3	+0.38519	−0.00714	+4.07542	+1.19119	−11.80211	−4.78465
Change 2% r_3 ray tracing	+0.02976	+0.00293	+0.67252	+0.09694	−1.18908	−0.62295
Change by new method	+0.031275	+0.00309	+0.71489	+0.099884	−1.19165	−0.64481
% discrepancy	+5.12	+5.5	+5.48	+2.01	+0.257	+3.51
Change 20% r_3 ray tracing	+0.20485	+0.02179	+4.28315	+0.61909	−6.68864	−4.77020
Change by new method	+0.31275	+0.03086	+7.1489	+0.98884	−11.9165	−6.4481
% discrepancy	+52.7	+41.6	+66.9	+59.7	+78.2	+35.2
			Condensed summary			
Discrepancy 2% in r_3	+5.12%	+5.5%	+5.48%	+2.01%	+0.257%	+3.51%
Discrepancy corrected	−0.08%	+0.3%	+0.28%	−3.19%	−4.94%	−1.69%
Discrepancy 20% in r_3	+52.7%	+41.6%	+66.9%	+59.7%	+78.2%	+35.2%
Discrepancy corrected	+0.7%	−10.4%	+14.9%	+7.7%	+26.2%	−16.8%

ceding group for quick and unerring inspection. Again we examine the new values of all aberrations for others that have become undesirably large. This process is repeated until the nearest approach to the desired system has been attained.

Form No. V offers the opportunity to determine the detailed over-all performance of the system, and to quickly determine the best that can be accomplished with the system as set up. An experienced man can also glean from these results what type of additional elements are required and where they should be placed in order to further improve the system.

QUANTITATIVE ANALYSIS

Since the differential adjustment method here presented, made use of the left-hand digits (43.2796) of the ray tracing data, it was at first intended to attempt a quantitative test of the method by using a slide rule, but to avoid having to recompute in case four places turned out to be insufficient, it was finally decided to use seven places. After much disappointing effort, it became evident that seven places were not enough to clear the region to be studied because, no doubt, of accumulated errors of ray tracing. In the case of the lens here chosen for study, the Cooke anastigmat, series IV, as many as three significant figures were lost in the ray tracing procedure. Sometimes this situation can be ameliorated by making slight changes in either the conditions of the test or the prescription. Sometimes other tricks are possible, but it did not seem wise to attempt a correction of this kind at this time. At any rate, the errors of ray tracing had accumulated to such an extent by the time the final image space was reached that the changes in aberrations, and in some cases the aberrations themselves (those which were already small) were completely obliterated by ray tracing errors.

It might be of interest to the reader how these errors creep into ray tracing data. There were a number of more or less serious cases encountered in this test, the worst of which occurred in the marginal ray at the 4th surface. The following appears at that point in the ray tracing data.

Item	Quantity	Angle
15	U	10° 42′ 27.89″
16	I	17° 14′ 59.96″
17	$I+U$	27° 57′ 27.85″
18	$-I'$	−27° 51′ 2.99″
19	U'	0° 6′ 24.86″

Item 18 comes from a seven place number above (Item 10, Form No. I). It is therefore in doubt in the 7th place, and fixes the error as a few tenths of a second. This absolute error is passed on to item 19, but there it represents an error in the 4th place, a loss of precision of three places.

In order to see what could be done with a bad situation, it seemed more desirable to retrace the rays to a greater degree of precision. Ten places were chosen. This moved the ray tracing errors beyond the region where the aberrations and

their variations occurred and left only the discrepancies resulting from the curvature of the differentiated function in the region. It thus became clear in this case that, in order to check the differential adjustment method against ray tracings, it was necessary to carry ray tracings to ten places in order to check the differential adjustment coefficients obtained from five places. The check then showed clearly the increasing discrepancy with increasing size of the change in the parameter, and showed that the percent of discrepancy between the change in an aberration calculated from ray tracing data and the change in the aberration calculated with the aid of the adjustment coefficients was of the general order of one-half the percent change in the parameter times the number of surfaces to be passed before reaching the final image including the one where the change in parameter occurs. A correcting equation which is a little more nearly right in this particular case, at least, is,

$$\%\ \text{change in aberration} = 0.65 \times \text{number of surfaces} \times \text{percent change in the parameter.}$$

The aberrations as calculated by the adjustment method are always numerically too large.

Table II shows the results obtained by a decrease of 2 percent and 20 percent in r_3, the radius of curvature of the third surface, counted in the direction of ray tracing.

The values appearing in this table have been carried to the full precision of the data as it was obtained, not because the data would ever be used this far out, but to show to what extent precision is lost during the progress of the work, and to obtain usable figures in the percent discrepancies. For instance, had I started with only three digits, I would have ended up with a wholly incorrect value for the percent discrepancy in the case of transverse chromatic (see Table II), as errors in the third place would have changed the first digit in 5.5 percent recorded for this aberration.

It will now be apparent that an aberration calculated from ray tracings, might be quite incorrect and, since the ten places required to clear the calculations of ray tracing errors of the order of magnitude of the aberrations under examination, is far beyond the practice of the glass shop, and, since the calculations, carried only to the point to which the glass shop can work, leaves much uncertainty as to the absolute value of those aberrations at times, it follows that attempts to correct the optical system in such cases by starting from these doubtful aberrations mean little. Corrections thus made, through the use of the method herein described, or any other method, upon such doubtful values could conceivably lead to an impairment and not an improvement in the optical system. Corrections made through the use of the new method on values of aberrations experimentally measured on a sample system made up according to the prescription used in the ray tracings from which the differential adjustment coefficients were obtained would accomplish much more than the customary procedure.

The Reflection and Transmission of Light by Any System of Parallel Isotropic Films

A. W. CROOK
Physics Department, University College, Southampton, England
(Received July 2, 1948)

By consideration of the formulae for the reflection and transmission of up to three films, it has been possible to form a generalization for an arbitrary number of films. This generalization is stated as a simple rule whereby the expressions for the reflected and transmitted amplitudes can be readily formed in terms of the Fresnel coefficients of the various boundaries. The generalization is proved by induction. It is also shown that the solutions of the problem by the consideration of the electromagnetic disturbance in each medium, and by the summation of multiple beams, lead to the same result.

INTRODUCTION

A NUMBER of authors have given expressions for the reflectivity and transmissivity of a finite number of films,[1–14] and some have indicated how, by the continuation of the processes they have employed, the results may be extended to any number of films. Mooney[15] and Caballero[16] have considered the case of normal incidence. Three recent papers have considered the problem in its most general form. Weinstein[17] and Herpin[18] have tackled the problem by the consideration of the electromagnetic disturbance in each medium and give expressions for the reflectivity and transmissivity in terms of matrices. A. Vašiček[19] has considered the problem from the point of view of the multiplicity of reflected and transmitted beams in the system. To deal with the complexity into which this leads he introduces what he terms "basic amplitudes." The results obtained in this paper differ from those of Vašiček, and it is suggested that Vašiček's assumption of "basic amplitudes" is incorrect.

[1] Lord Rayleigh, *Scientific Papers* (Cambridge University Press, Teddington, England, 1902), Vol. 3, p. 63.
[2] P. Drude, *Theory of Optics* (Longmans Green and Company, New York, 1902), p. 302.
[3] T. C. Fry, J. Opt. Soc. Am. **16**, 1 (1928).
[4] J. B. Nathanson and C. L. Bartberger, J. Opt. Soc. Am. **29**, 417 (1939).
[5] C. E. Leberknight and B. Lustman, J. Opt. Soc. Am. **29**, 59 (1939).
[6] K. B. Blodgett, Phys. Rev. **57**, 921 (1940).
[7] J. A. Stratton, *Electro Magnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 511.
[8] K. M. Greenland, Nature **152**, 290 (1943).
[9] A. B. Winterbottom, Trans. Faraday Soc. **42**, 478 (1946).
[10] A. Vašiček, J. Opt. Soc. Am. **37**, 145 (1947).
[11] P. King and L. B. Lockhart, J. Opt. Soc. Am. **36**, 513 (1946).
[12] R. Messner, Optik **2**, 228 (1947).
[13] L. B. Lockhart and P. King, J. Opt. Soc. Am. **36**, 689 (1947).
[14] L. N. Hadley and D. M. Dennison, J. Opt. Soc. Am. **37**, 451 (1947).
[15] R. L. Mooney, J. Opt. Soc. Am. **35**, 574 (1945).
[16] D. L. Caballero, J. Opt. Soc. Am. **37**, 176 (1947).
[17] W. Weinstein, J. Opt. Soc. Am. **37**, 576 (1947).
[18] A. Herpin, Comptes Rendus **225**, 182 (1947).
[19] A. Vašiček, J. Opt. Soc. Am. **37**, 623 (1947).

SOLUTION BY THE CONSIDERATION OF THE ELECTROMAGNETIC DISTURBANCE

The Electromagnetic Disturbance in Each Film

In Fig. 1 are shown n films sectioned in a plane $y=$const., with the various media numbered from 1 to $n+2$. Let the plane of incidence be $y=$const. and let the disturbance in the films on a plane $x=$const. be considered. Then at each point in this plane the disturbance consists of a wave moving downwards towards the right and a wave moving upwards to the right, both being equally inclined to a line parallel to the z axis. These waves in the kth medium can be represented by

$$\xi=\xi_0 \exp i\omega[t-\mu_k(a_k x-c_k z)/v] \qquad (1.1)$$

for the wave moving downwards, and by

$$\psi=\psi_0 \exp i\omega[t-\mu_k(a_k x+c_k z)/v] \qquad (1.2)$$

for the wave moving upwards, where ξ and ψ may be electric or magnetic vectors, v is the velocity of light in vacuum, a and c are direction cosines, and μ_k is the refractive index of the medium. This treatment follows Drude[2] (p. 303). By Snell's law

$$\mu_1 a_1=\mu_2 a_2=\cdots=\mu_k a_k=\cdots=\mu_{n+2} a_{n+2}, \qquad (1.3)$$

and since a and c are direction cosines

$$c_k=+(1-a_k^2)^{\frac{1}{2}}=+\{1-(\mu_1 a_1/\mu_k)^2\}^{\frac{1}{2}}, \qquad (1.4)$$

the positive sign being taken because of the

choice of coordinates in Fig. 1. In these equations ξ_0, ψ_0, μ_k, a_k, and c_k may be complex. Equations (1.1) and (1.2) give the relationship between R and P, S, and Q in each film. If the thickness of the kth medium is d_k,

$$R_k = P_k \exp -i\phi_k, \tag{1.5}$$

$$S_k = Q_k \exp +i\phi_k, \tag{1.6}$$

where

$$\phi_k = -\omega\mu_k c_k d_k/v = -2\pi\mu_k c_k d_k/\lambda_0.$$

Assume that the first medium is non-absorbing, then by means of Eqs. (1.3) and (1.4) from the angle of incidence of the light in the first medium, each ϕ_k may be calculated. In the case of a non-absorbing medium ϕ_k will be real, in the case of an absorbing medium ϕ_k will be complex, and if $\mu_1 a_1/\mu_k = \mu_{k-1}a_{k-1}/\mu_k > 1$, i.e., total reflection has taken place at the $k-1$, k surface, ϕ_k will be a pure imaginary. In this case, however, since the sign of the imaginary must be such that the superficial wave is attenuated in the correct direction, i.e., as z decreases, (1.4) must be written

$$c_k = -i[(\mu_1 a_1/\mu_k)^2 - 1]^{\frac{1}{2}}. \tag{1.7}$$

(1.7) is in accordance with the expressions given in the standard works.[20–22]

The Boundary Conditions

According to Stokes the reflected amplitude, Q_1, and the transmitted amplitude, R_1, (Fig. 2) at the surface 1, 2 are given by

$$Q_1 = P_1 r_{12}, \quad R_2 = P_1 t_{12},$$

and

$$r_{12} = -r_{21}, \tag{2.1}$$

$$r_{12}^2 + t_{12}t_{21} = 1. \tag{2.2}$$

(A derivation of 2.1 and 2.2 is given by R. W. Wood in his *Physical Optics*,[21] p. 192.)

If the coefficients r and t are generalized so that they may be complex, (2.1) and (2.2) are satisfied by either set of Fresnel coefficients. If the incident electric vector is perpendicular to the plane of incidence,

$$r_{k\,k+1} = [\mu_k c_k - \mu_{k+1}c_{k+1}]/[\mu_k c_k + \mu_{k+1}c_{k+1}], \tag{2.3}$$

$$t_{k\,k+1} = 2\mu_k c_k/[\mu_k c_k + \mu_{k+1}c_{k+1}]. \tag{2.4}$$

If the incident electric vector is in the plane of incidence,

$$r_{k\,k+1} = [\mu_{k+1}c_k - \mu_k c_{k+1}]/[\mu_{k+1}c_k + \mu_k c_{k+1}], \tag{2.5}$$

$$t_{k\,k+1} = 2\mu_k c_k/[\mu_{k+1}c_k + \mu_k c_{k+1}]. \tag{2.6}$$

The Fresnel coefficients in their more usual form can be easily obtained from the above by putting

$$c_k = \cos\theta_k, \quad c_{k+1} = \cos\theta_{k+1},$$

and

$$\mu_{k+1}/\mu_k = \sin\theta_k/\sin\theta_{k+1},$$

where θ_k is the angle of incidence and θ_{k+1} the angle of refraction.

Since the Fresnel coefficients arise from the boundary conditions, if at each surface the waves are related by a treatment following Stokes, the boundary conditions are satisfied. Lord Rayleigh (see reference 1, p. 67) has stated that Stokes's treatment is inapplicable to the problem of a thin plate. His first objection appears to be answered by the electromagnetic theory. In answer to the second of his objections it will be shown later that if the films have vanishing thickness the equations give the correct reflection and transmission coefficients.

A Single Surface

At the surface $k-1$, k following Stokes,

$$R_k = P_{k-1}t_{k-1\,k} + S_k r_{k\,k-1},$$

$$Q_{k-1} = P_{k-1}r_{k-1\,k} + S_k t_{k\,k-1};$$

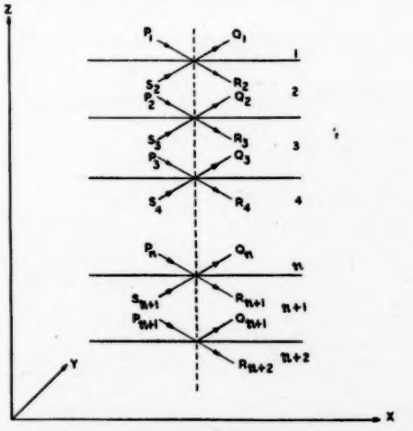


FIG. 1. The electromagnetic disturbance in a system of n films sectioned in a plane y = const., on a plane x = const.

[20] See reference 2, p. 296.
[21] R. W. Wood, *Physical Optics* (MacMillan Company, Ltd., London, 1939), third edition, p. 416.
[22] W. König, *Handbuch der Physik* 20, 229 (1928).

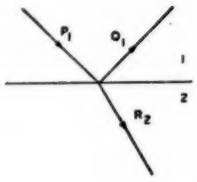


FIG. 2. Reflection and transmission at a single surface.

remembering (2.1) and (2.2), this is

$$P_{k-1}=[R_k+S_k r_{k-1\,k}]/t_{k-1\,k}, \tag{3.1}$$

$$Q_{k-1}=[R_k r_{k-1\,k}+S_k]/t_{k-1\,k}. \tag{3.2}$$

Application to the Problem of Multiple Films

Let there be n films. Then putting $k=n+2$ and $S_k=0$ in (3.1) and (3.2), P_{n+1} and Q_{n+1} can be expressed in terms of R_{n+2}. Putting $k=n+1$ in Eqs. (1.5) and (1.6), R_{n+1} and S_{n+1} can be expressed in terms of P_{n+1} and Q_{n+1}, and therefore in terms of R_{n+2}. Using (3.1) and (3.2), P_n and Q_n can be expressed in terms of R_{n+1} and S_{n+1}, and therefore in terms of R_{n+2}. In this manner, by starting with the transmitted wave and working upwards so that the incident and reflected waves are related to the transmitted wave, the process can be continued to give a solution of a problem of any number of films. The results are given for one, two, and three films.

One Film

$$Q_1/P_1=\alpha_{13}=[r_{12}+r_{23}\exp i2\phi_2]/[1+r_{12}r_{23}\exp i2\phi_2]. \tag{4.1}$$

Let $\alpha_{13}=N_{13}/D_{13}$, where N_{13} is the numerator of (4.1) and D_{13} the denominator. Then

$$R_3/P_1=\beta_{13}=t_{12}t_{23}\exp i\phi_2/D_{13}. \tag{4.2}$$

Two Films

$$Q_1/P_1=\alpha_{14}=\left[\frac{r_{12}+r_{23}\exp i(2\phi_2)+r_{34}\exp i(2\phi_2+2\phi_3)+r_{12}r_{23}r_{34}\exp i(2\phi_3)}{1+r_{12}r_{23}\exp i(2\phi_2)+r_{12}r_{34}\exp i(2\phi_2+2\phi_3)+r_{23}r_{34}\exp i(2\phi_3)}\right], \tag{4.3}$$

$$R_4/P_1=\beta_{14}=t_{12}t_{23}t_{34}\exp i(\phi_2+\phi_3)/D_{14}. \tag{4.4}$$

Three Films

$$\alpha_{15}=\left[\frac{\begin{array}{r}r_{12}+r_{23}\exp i(2\phi_2)+r_{34}\exp i(2\phi_2+2\phi_3)+r_{45}\exp i(2\phi_2+2\phi_3+2\phi_4)+r_{12}r_{23}r_{34}\exp i(2\phi_3)\\ +r_{12}r_{34}r_{45}\exp i(2\phi_4)+r_{12}r_{23}r_{45}\exp i(2\phi_3+2\phi_4)+r_{23}r_{34}r_{45}\exp i(2\phi_2+2\phi_4)\end{array}}{\begin{array}{r}1+r_{12}r_{23}\exp i(2\phi_2)+r_{12}r_{34}\exp i(2\phi_2+2\phi_3)+r_{12}r_{45}\exp i(2\phi_2+2\phi_3+2\phi_4)+r_{23}r_{34}\exp i(2\phi_3)\\ +r_{34}r_{45}\exp i(2\phi_4)+r_{23}r_{45}\exp i(2\phi_3+2\phi_4)+r_{12}r_{23}r_{34}r_{45}\exp i(2\phi_2+2\phi_4)\end{array}}\right], \tag{4.5}$$

$$\beta_{15}=t_{12}t_{23}t_{34}t_{45}\exp i(\phi_2+\phi_3+\phi_4)/D_{15}, \tag{4.6}$$

where the αs and βs will, in general, be complex.

THE GENERALIZATION FOR AN ARBITRARY NUMBER OF FILMS

The Statement of the Generalization

Set down the scheme

$$\begin{array}{l} r_{12}\\ r_{23}\exp i(2\phi_2)\\ r_{34}\exp i(2\phi_2+2\phi_3)\\ \downarrow\\ r_{k\,k+1}\exp i(2\phi_2+2\phi_3+\cdots+2\phi_k)\\ \downarrow\\ r_{n\,n+1}\exp i(2\phi_2+2\phi_3+\cdots+2\phi_n)\\ r_{n+1\,n+2}\exp i(2\phi_2+2\phi_3+\cdots+2\phi_n+2\phi_{n+1}).\end{array}$$

Then let

$$r_{l\,l+1}r_{k\,k+1}r_{p\,p+1}\exp i\{+(2\phi_2+\cdots+2\phi_l)-(2\phi_2+\cdots+2\phi_k)+(2\phi_2+\cdots+2\phi_p)\},$$

where $l<k<p$ be called a third product with the phase pattern $+-+\cdots$, and let this be an example of an odd product with a phase pattern starting with $+$. Then $N_{1\,n+2}$ is given by the sum

$$r_{12}+r_{23}\exp i(2\phi_2)+\cdots+r_{n+1\,n+2}\exp i(2\phi_2+2\phi_3+\cdots+2\phi_{n+1})$$

plus

the sum of all the possible third products with phase pattern $+-+$

plus

the sum of all the possible fifth products with phase pattern $+-+-+$

plus

up to nth products if n is odd and an $(n+1)$th product if n is even. So $N_{1\,n+2}$ can be said to be

the sum of all the possible odd products with alternating phase patterns starting with $+$. $D_{1\,n+2}$ is given by the sum

1

plus

all the possible second products with phase pattern $-+$

plus

all the possible fourth products with phase pattern $-+-+$

plus

up to nth products if n is even and an $(n+1)$th product if n is odd. So $D_{1\,n+2}$ is the sum of one plus all the possible even products with alternating phase patterns starting with $-$. Hence $\alpha_{1\,n+2}$ can be written down. Similarly, $\beta_{1\,n+2}$ is given by

$$\beta_{1\,n+2}=t_{12}t_{23}\cdots t_{n+1\,n+2}\times\exp i(\phi_2+\phi_3+\cdots+\phi_{n+1})/D_{1\,n+2}.$$

As an example of these rules, their application to a three-film problem is given. The scheme is

$$r_{12}$$
$$r_{23}\exp i(2\phi_2)$$
$$r_{34}\exp i(2\phi_2+2\phi_3)$$
$$r_{45}\exp i(2\phi_2+2\phi_3+2\phi_4).$$

The numerator is

$$r_{12}+r_{23}\exp i(2\phi_2)+r_{34}\exp i(2\phi_2+2\phi_3)+r_{45}\exp i(2\phi_2+2\phi_3+2\phi_4)$$

plus

$$r_{12}r_{23}r_{34}\exp i(-2\phi_2+2\phi_2+2\phi_3)+r_{12}r_{34}r_{45}\exp i(-2\phi_2-2\phi_3+2\phi_2+2\phi_3+2\phi_4)+r_{12}r_{23}r_{45}\exp i(-2\phi_2+2\phi_2+2\phi_3+2\phi_4)+r_{23}r_{34}r_{45}\exp i(+2\phi_2-2\phi_2-2\phi_3+2\phi_2+2\phi_3+2\phi_4),$$

and the denominator is

1

plus

$$r_{12}r_{23}\exp i(2\phi_2)+r_{12}r_{34}\exp i(2\phi_2+2\phi_3)+r_{12}r_{45}\exp i(2\phi_2+2\phi_3+2\phi_4)+r_{23}r_{34}\exp i(-2\phi_2+2\phi_2+2\phi_3)+r_{34}r_{45}\exp i(-2\phi_2-2\phi_3+2\phi_2+2\phi_3+2\phi_4)+r_{23}r_{45}\exp i(-2\phi_2+2\phi_2+2\phi_3+2\phi_4)$$

plus

$$r_{12}r_{23}r_{34}r_{45}\exp i(+2\phi_2-2\phi_2-2\phi_3+2\phi_2+2\phi_3+2\phi_4).$$

These expressions for the numerator and denominator agree with (4.5). The cases of 1 and 2 films are also in agreement with (4.1) and (4.3), respectively.

The Proof of the Generalization

Consider a system of $(n+1)$ films as shown in Fig. 3. Then by (1.5), (1.6), (3.1), and (3.2)

$$P_1=[P_2\exp-i\phi_2+r_{12}Q_2\exp i\phi_2]/t_{12}, \quad (5.1)$$

$$Q_1=[r_{12}P_2\exp-i\phi_2+Q_2\exp i\phi_2]/t_{12}. \quad (5.2)$$

So that

$$Q_1/P_1=[r_{12}+(Q_2/P_2)\exp i2\phi_2]/[1+r_{12}(Q_2/P_2)\exp i2\phi_2]. \quad (5.3)$$

Now Q_2/P_2 is given by the application of the generalization to the n films between media 2 and $(n+3)$. Writing

$$Q_2/P_2 \quad \text{as} \quad N_{2\,n+3}/D_{2\,n+3},$$

$$Q_1/P_1=[r_{12}D_{2\,n+3}+N_{2\,n+3}\exp i2\phi_2]/[D_{2\,n+3}+r_{12}N_{2\,n+3}\exp i2\phi_2]. \quad (5.4)$$

In Fig. 4 are shown the schemes for n and $n+1$ films. In this the n films are derived from the $n+1$ films by the removal of the first of the $n+1$ films so that the first medium in the n film case is medium 2. It is convenient to note now that an element in the scheme for $n+1$ films is the element on the same line for n films multiplied by $\exp i2\phi_2$, except in the case of r_{12}.

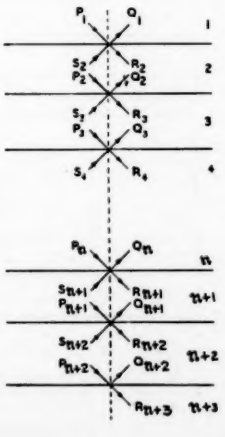


FIG. 3. The electromagnetic disturbance in a system of $n+1$ films.

The Numerator of (5.4)

The Product $r_{12}D_{2\,n+3}$

$D_{2\,n+3}$ contains all the possible even products drawn from the elements r_{23} to $r_{n+2\,n+3}\times\exp i(2\phi_3+\cdots+2\phi_{n+2})$ with the phase patterns $-+-+\cdots$. Since $D_{2\,n+3}$ contains even products, the phase patterns contain as many minus signs as they do positive signs, so that multiplying each element from which $D_{2\,n+3}$ is constructed by $\exp i2\phi_2$ does not alter the value of $D_{2\,n+3}$. $D_{2\,n+3}$ may therefore be considered as constructed from the elements $r_{23}\exp i(2\phi_2)$ to $r_{n+2\,n+3}\times\exp i(2\phi_2+\cdots+2\phi_{n+2})$ in the $(n+1)$ column of Fig. 4. Therefore, multiplying $D_{2\,n+3}$ by r_{12} gives the sum of all the odd products with phase patterns $+-+\cdots$ which contain r_{12} that can be obtained from the scheme for $(n+1)$ films.

The Product $N_{2\,n+3}\exp i(2\phi_2)$

$N_{2\,n+3}$ contains all the possible odd products drawn from r_{23} to $r_{n+2\,n+3}\exp i(2\phi_3+\cdots+2\phi_{n+2})$ with the phase pattern $+-+\cdots$. If each element from which $N_{2\,n+3}$ is constructed is multiplied by $\exp i(2\phi_2)$, since each phase pattern contains one more positive sign than negative signs, this is equivalent to multiplying $N_{2\,n+3}$ by $\exp i2\phi_2$. But multiplying each element by $\exp i2\phi_2$ gives the corresponding elements in the $(n+1)$ column. So $N_{2\,n+3}\exp i(2\phi_2)$ gives the sum of all the odd products with the phase pattern $+-+\cdots$, which do not contain r_{12}, that can be obtained from the scheme for $(n+1)$ films. Therefore

$$r_{12}D_{2\,n+3}+N_{2\,n+3}\exp i(2\phi_2)=N_{1\,n+3}. \quad (5.5)$$

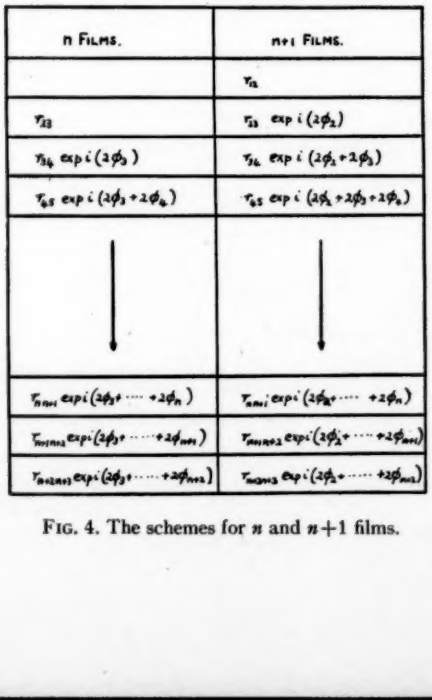


FIG. 4. The schemes for n and $n+1$ films.

The Denominator of (5.4)

$D_{2\,n+3}$

As shown in the consideration of the product $r_{12}D_{2\,n+3}$, $D_{2\,n+3}$ gives the sum of all the even products which do not include r_{12} that can be constructed from the scheme for $(n+1)$ films.

The Product $r_{12}N_{2\,n+3}\exp i(2\phi_2)$

As shown in the consideration of the product $N_{2\,n+3}\exp i(2\phi_2)$, this product contains all the odd products with phase pattern $+-+\cdots$ which do not contain r_{12} that may be constructed from the elements $r_{23}\exp i(2\phi_2)$ to $r_{n+2\,n+3}\exp i(2\phi_2+\cdots+2\phi_{n+2})$. Therefore, multiplying by r_{12} gives the sum of all the even products which do contain r_{12} with the phase pattern $-+-+\cdots$ that may be constructed from the elements in the scheme for $(n+1)$ films. Therefore

$$D_{2\,n+3}+r_{12}N_{2\,n+3}\exp i(2\phi_2)=D_{1\,n+3}. \quad (5.6)$$

Equations (5.5) and (5.6) prove that if the generalization is true for n films, it is also true for $n+1$ films. Since it is true for α_{13}, α_{14}, and α_{15} it must be true for $\alpha_{1\,n+2}$.

The Transmitted Wave

Expressing P_2 and Q_2 in terms of R_{n+3} by means of the statement of the generalizations for the transmitted and reflected waves, and rearranging, (5.1) becomes

$$R_{n+3}/P_1=t_{12}t_{23}\cdots t_{n+2\,n+3}\exp i(\phi_2+\cdots+\phi_{n+2})/\{D_{2\,n+3}+r_{12}N_{2\,n+3}\exp i(2\phi_2)\},$$

and by (5.6) this is

$$R_{n+3}/P_1=\beta_{1\,n+3}=t_{12}t_{23}\cdots t_{n+2\,n+3}\times\exp i(\phi_2+\cdots+\phi_{n+2})/D_{1\,n+3}. \quad (5.7)$$

This proves that if the generalization for the transmitted ray is true for n films, it is true also for $n+1$ films. Since it is true for β_{13}, β_{14}, and β_{15}, it must be true for $\beta_{1\,n+2}$.

The Number of Terms in the Numerator and Denominator

The Numerator

If there are n films there are $n+1$ elements in the scheme, so the number of possible third products is the number of ways in which 3 things can be selected from $n+1$ things. Similarly for the other products, so the number of terms in the numerator is

$$^{n+1}C_1+{}^{n+1}C_3+{}^{n+1}C_5+\cdots,$$

ending in $^{n+1}C_n$ if n is odd and $^{n+1}C_{n+1}$ if n is even.

The Denominator

The number of terms is

$$^{n+1}C_0+{}^{n+1}C_2+{}^{n+1}C_4+\cdots,$$

ending in $^{n+1}C_n$ if n is even and $^{n+1}C_{n+1}$ if n is odd. By putting $x=-1$ in the expansion of $(1+x)^{n+1}$ it is shown that the number of terms in the numerator is equal to the number of terms in the denominator, and if $x=1$ it is seen that both are 2^n. (5.5) and (5.6) show immediately that if one film is added there are 2^{n+1} terms in the resulting numerator and denominator.

A Relationship between the Numerator and the Denominator

If all the terms in the numerator which contain r_{12} are divided by r_{12}, the result is all the even products with the correct phase pattern which do not contain r_{12}. If all the terms in the numerator which do not contain r_{12} are multiplied by r_{12}, the result is all the even products which do contain r_{12} with the correct phase pattern. By noticing this, the labor of writing down the expressions is reduced. The relationship may be expressed formally by

$$\alpha_{1\,n+2}=[X+Yr_{12}]/[Xr_{12}+Y], \qquad (5.8)$$

where X and Y do not contain r_{12}.

SOLUTION BY THE SUMMATION OF MULTIPLE BEAMS

Most elementary treatments of the optical properties of a thin film deal with the problem by the consideration of the multiply reflected beams in the film, for example, in the theory of Newton's rings. This method is applicable to the problem of an arbitrary number of films, and it will now be shown that it leads to the results already obtained.

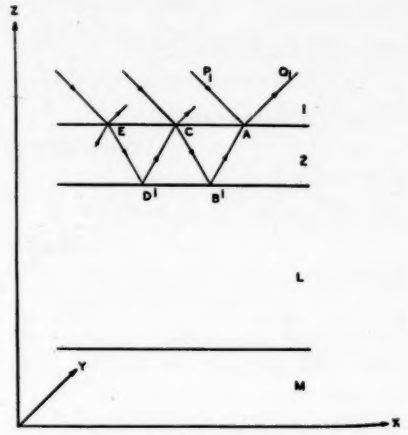


FIG. 5. The beams to be summed in calculating the reflected wave.

The Reflected Wave

In Fig. 5 is shown a film of medium 2 on top of a system of films denoted by L which rest on a base of medium M. The wave Q_1 reflected at A is due to

the wave incident at A, reflected at A,

plus

the wave incident at C, transmitted at C, reflected at B', transmitted at A,

plus

the wave at E giving rise by multiple reflection to a disturbance at A,

and so on for an infinity of points to the left of E. Noting that at each reflection the wave in medium 2 changes the sign of its z direction cosine and that the wave in medium 2 is to be related always to the wave in medium 1 at the point in the 1,2 surface which is in the same plane $x=$ const., so that since $\mu_1 a_1=\mu_2 a_2$, no phase term including the x direction cosine, is introduced due to the progress of this plane in the positive direction of the x axis, the sum of the

multiple beams is

$$Q_1 = P_1[r_{12} + t_{12}\alpha_{2M}t_{21}\exp i(2\phi_2) + t_{12}\alpha_{2M}r_{21}\alpha_{2M}t_{21}\exp i(4\phi_2) + \cdots]$$
$$= P_1[r_{12} + (t_{12}\alpha_{2M}t_{21}\exp i2\phi_2)/(1 - \alpha_{2M}r_{21}\exp i2\phi_2)].$$

By (2.1) and (2.2) this becomes

$$\alpha_{1M} = Q_1/P_1 = [r_{12} + \alpha_{2M}\exp i(2\phi_2)]/[1 + r_{12}\alpha_{2M}\exp i(2\phi_2)]. \quad (6.1)$$

The one film problem is dealt with by substituting r_{23} for α_{2M} in (6.1). The two-film problem is dealt with by substituting from the solution of the one film problem, the value of α_{24} for α_{2M} in (6.1) and so on. If this is done the results are the same as (4.1), (4.3), and (4.5) and can be made the subject of the same generalization. If $\alpha_{2M} = \alpha_{2\,n+3}$, (6.1) becomes

$$\alpha_{1\,n+3} = [r_{12}D_{2\,n+3} + N_{2\,n+3}\exp i(2\phi_2)]/[D_{2\,n+3} + r_{12}N_{2\,n+3}\exp i(2\phi_2)],$$

which is the same as (5.4). From this the proof by induction is the same as that given previously.

The Transmitted Wave

In Fig. 6 are shown the beams which have to be summed. An argument similar to that used in the case of the reflected wave leads to the result

$$P_2 = P_1[t_{12}\exp i(\phi_2) + t_{12}\alpha_{2M}r_{21}\exp i(3\phi_2) + t_{12}\alpha_{2M}r_{21}\alpha_{2M}r_{21}\exp i(5\phi_2) + \cdots].$$

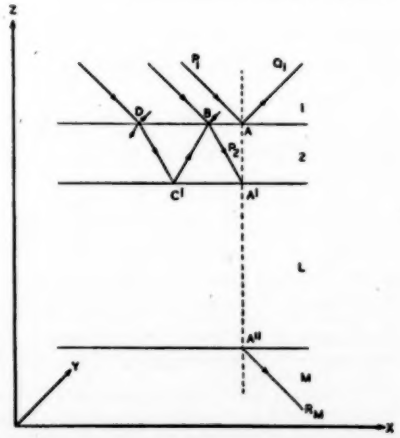


FIG. 6. The beams to be summed in calculating the transmitted wave.

This is

$$P_2 = P_1t_{12}\exp i(\phi_2)/[1 + r_{12}\alpha_{2M}\exp i(2\phi_2)],$$

giving the relationship between the wave incident at A and the wave incident at A'.

The wave R_M transmitted at A'' is

$$R_M = P_2\beta_{2M} = P_1\beta_{2M}t_{12}\exp i(\phi_2)/[1 + r_{12}\alpha_{2M}\exp i(2\phi_2)],$$

or

$$\beta_{1M} = R_M/P_1 = \beta_{2M}D_{2M}t_{12}\exp i(\phi_2)/[D_{2M} + r_{12}N_{2M}\exp i(2\phi_2)]. \quad (6.2)$$

The one-film problem is dealt with by replacing the subscript $2M$ by the subscript 23 in (6.2), the two-film problem by substituting in (6.2) the values of β_{24} and D_{24} and N_{24} obtained from the solution of the one-film problem, and so on. This process leads to equations identical with (4.2), (4.4), and (4.6), so that the same generalization can be made. If in (6.2) the subscript $2M$ is replaced by the subscript $2n+3$ and the generalized expression for $\beta_{2\,n+3}$ employed, (6.2) becomes

$$\beta_{1\,n+3} = t_{12}t_{23}\cdots t_{n+2\,n+3} \times \exp i(\phi_2 + \phi_3 + \cdots + \phi_{n+2})/D_{1\,n+3},$$

which is the same as (5.7). This treatment by the summation of multiple beams therefore leads to the same results as previously obtained.

AN ENERGY RELATIONSHIP

As shown by Lord Rayleigh (see reference 1, p. 67) a film on a perfectly reflecting base can show no interference colors. If the base on which the film system is deposited is a perfect reflector and all the films are non-absorbing, all the incident energy must be reflected so that $|\alpha|^2 = 1$ irrespective of the number and thicknesses of the films. However, to test the generalization in the most general way, a number of cases have to be considered. For the above consideration to apply it is not necessary for the perfectly reflecting surface to be the last surface in the system. Let the perfectly reflecting surface be the surface k, $k+1$. First imagine that the perfect reflection arises from the $(k+1)$th medium, optically a perfect metal; i.e., putting $\mu_{k+1} = \nu_{k+1} \times (1 - i\kappa_{k+1})$, the absorption coefficient κ_{k+1} is infinite; then $r_{k\,k+1} = +1$ if the electric vector is in the plane of incidence and $r_{k\,k+1} = -1$ if the electric vector is perpendicular to the plane of

incidence; also $\exp i\phi_{k+1}=0$, since if κ_{k+1} is infinite

$$\exp i\phi_{k+1}=[\exp -i\omega\nu_{k+1}c_{k+1}d_{k+1}/v]\times[\exp -\omega\nu_{k+1}\kappa_{k+1}c_{k+1}d_{k+1}/v]=0.$$

Under these circumstances, if all the media above the k, $k+1$ surface are non-absorbing, $|\alpha|^2$ should be unity irrespective of their number and thicknesses, and this should also be true irrespective of whether the media below the $k+1$, $k+2$ surface are absorbing or not. If the generalization is correct it should therefore be possible to prove that if the $(k+1)$th medium is optically a perfect metal, all the media above non-absorbing and the media below of any nature whatsoever, that $|\alpha|^2=1$. Secondly, consider the perfect reflection of the k, $k+1$ surface resulting from total internal reflection. The generalization depends upon the propagation equation, Snell's law, and the boundary conditions, all of which remain valid at total reflection, so total reflection does not make the generalization inapplicable. Consider total reflection occurring at the k, $k+1$ surface. Let all the media above be non-absorbing; then it might appear that $|\alpha|^2$ should be unity irrespective of the media below the $k+1$, $k+2$ surface. This will be true, however, only if the disturbance in the $(k+1)$th medium is so attenuated that there is a negligible disturbance at the $k+1$, $k+2$ surface. If this is not so, energy will pass through the less dense medium and will be reflected back in part by the succeeding films, and the reflectivity of the system will depend upon the constants of all the films and all the media. $|\alpha|^2$ should, however, be unity if the k, $k+1$ surface is the last surface in the system. It will now be shown that these results, which are physically necessary, can be obtained from the generalization.

Write down the scheme for n films.

$$\begin{array}{l} r_{12} \\ \downarrow \\ r_{k-1\,k}\exp i(2\phi_2+\cdots+2\phi_{k-1}) \\ r_{k\,k+1}\exp i(2\phi_2+\cdots+2\phi_{k-1}+2\phi_k) \\ r_{k+1\,k+2}\exp i(2\phi_2+\cdots+2\phi_{k-1}+2\phi_k+2\phi_{k+1}) \\ \downarrow \\ r_{n\,n+1}\exp i(2\phi_2+\cdots+2\phi_n) \\ r_{n+1\,n+2}\exp i(2\phi_2+\cdots+2\phi_n+2\phi_{n+1}). \end{array}$$

Multiply this scheme by $\exp -i(2\phi_2+\cdots+2\phi_k)$ and obtain scheme two.

$$\begin{array}{l} r_{12}\exp -i(2\phi_2+\cdots+2\phi_{k-1}+2\phi_k) \\ \downarrow \\ r_{k-1\,k}\exp -i(2\phi_k) \\ r_{k\,k+1} \\ r_{k+1\,k+2}\exp i(2\phi_{k+1}) \\ \downarrow \\ r_{n\,n+1}\exp i(2\phi_{k+1}+\cdots+2\phi_n) \\ r_{n+1\,n+2}\exp i(2\phi_{k+1}+\cdots+2\phi_n+2\phi_{n+1}). \end{array}$$

Let the numerator drawn from scheme two be $N_{1\,n+2}'$ and the denominator be $D_{1\,n+2}'$. Then, since the phase patterns of the terms in a numerator contain one more positive sign than negative signs,

$$N_{1\,n+2}'=N_{1\,n+2}\exp -i\times(2\phi_2+\cdots+2\phi_{k-1}+2\phi_k) \quad (7.1)$$

and

$$D_{1\,n+2}'=D_{1\,n+2}, \quad (7.2)$$

since the phase patterns of the terms in a denominator contain as many positive signs as negative signs. Let X_1 denote the sum of one plus the even products with phase pattern $-+-+\cdots$, and Y_1 the sum of the odd products with phase pattern $+-+\cdots$ drawn from the elements above $r_{k\,k+1}$ in scheme two. Note that the complex conjugate of an expression of the form

$$x_1\exp i\theta_1+x_2\exp i\theta_2+\cdots+x_m\exp i\theta_m$$

is $x_1\exp -i\theta_1+x_2\exp -i\theta_2+\cdots+x_m\exp -i\theta_m$ if all the xs and θs are real. Then since all the media above the k, $k+1$ surface are non-absorbing and no total reflection is to occur until the k, $k+1$ surface, r_{12} to $r_{k-1\,k}$ and ϕ_2 to ϕ_k are real so that $\bar{X}_1$ the complex conjugate of X_1 is the sum of one plus all the even products with phase pattern $+-+-\cdots$ and $\bar{Y}_1$ is the sum of all the odd products with phase pattern $-+-\cdots$ drawn from the elements of scheme two above the element $r_{k\,k+1}$. Let X_2 denote the sum of one plus the even products with phase pattern $-+\cdots$ and Y_2 the sum of the odd products with phase pattern $+-+\cdots$ drawn from the elements below $r_{k\,k+1}$ in scheme two. The complex conjugates of X_2 and Y_2 will not necessarily bear a similar relationship to X_2 and Y_2, as did $\bar{X}_1$ and $\bar{Y}_1$ to X_1 and Y_1, since $r_{k+1\,k+2}$ to $r_{n+1\,n+2}$ and ϕ_{k+1} to ϕ_{n+1} may be complex. Then the odd products with the phase pattern $+-+\cdots$ drawn from scheme two which do not contain

$r_{k\,k+1}$ are given by

$$\bar{X}_1Y_2+Y_1X_2,$$

and the odd products which do contain $r_{k\,k+1}$ by

$$r_{k\,k+1}(\bar{X}_1X_2+Y_1Y_2).$$

So

$$N_{1\,n+2}'=[(\bar{X}_1Y_2+Y_1X_2)+r_{k\,k+1}(\bar{X}_1X_2+Y_1Y_2)]. \quad (7.3)$$

By a similar argument it is found that

$$D_{1\,n+2}'=[(X_1X_2+\bar{Y}_1Y_2)+r_{k\,k+1}(X_1Y_2+\bar{Y}_1X_2)]. \quad (7.4)$$

By (7.1) and (7.2)

$$|N_{1\,n+2}'/D_{1\,n+2}'|^2=|N_{1\,n+2}/D_{1\,n+2}|^2=|\alpha_{1\,n+2}|^2.$$

So

$$|\alpha_{1\,n+2}|^2=\frac{[(\bar{X}_1Y_2+Y_1X_2)+r_{k\,k+1}(\bar{X}_1X_2+Y_1Y_2)][(X_1\bar{Y}_2+\bar{Y}_1\bar{X}_2)+\bar{r}_{k\,k+1}(X_1\bar{X}_2+\bar{Y}_1\bar{Y}_2)]}{[(X_1X_2+\bar{Y}_1Y_2)+r_{k\,k+1}(X_1Y_2+\bar{Y}_1X_2)][(\bar{X}_1\bar{X}_2+Y_1\bar{Y}_2)+\bar{r}_{k\,k+1}(\bar{X}_1\bar{Y}_2+Y_1\bar{X}_2)]}, \quad (7.5)$$

putting $|r_{k\,k+1}|^2=1$, which is true for the cases of reflection at a perfect metal and of total reflection; and by multiplying out the numerator and denominator of (7.5) it can be shown that $|\alpha_{1\,n+2}|^2$ is of the form

$$|\alpha_{1\,n+2}|^2=[A+Br_{k\,k+1}+\bar{B}\bar{r}_{k\,k+1}]/[A+Cr_{k\,k+1}+\bar{C}\bar{r}_{k\,k+1}], \quad (7.6)$$

where $B+\bar{B}=C+\bar{C}$ but $B\neq C$. In the case of reflection at a perfect metal $\exp i\phi_{k+1}=0$, so that $X_2=1$, $Y_2=0$ and (7.5) reduces to

$$|\alpha_{1\,n+2}|^2=\{[Y_1+r_{k\,k+1}\bar{X}_1][\bar{Y}_1+\bar{r}_{k\,k+1}X_1]\}/\{[X_1+r_{k\,k+1}\bar{Y}_1][\bar{X}_1+\bar{r}_{k\,k+1}Y_1]\}. \quad (7.7)$$

In the case of total reflection if the magnitude of $\omega\mu_{k+1}c_{k+1}d_{k+1}/v$ is large where c_{k+1} is obtained from (1.7), $\exp i\phi_{k+1}=0$ and (7.5) reduces to (7.7). If, however, $\exp i\phi_{k+1}$ cannot be considered to be zero, as at total reflection $r_{k\,k+1}$ is complex, from (7.6) it can be seen that $|\alpha_{1\,n+2}|^2$ is a function of all the rs and ϕs as required by the physical considerations. If $[(\mu_1a_1/\mu_{k+1})^2-1]=0$, c_{k+1} is zero, $r_{k\,k+1}$ is $+1$, and $r_{k+1\,k+2}$ is -1. If c_{k+1} is zero, $\exp i\phi_{k+1}$ is unity and X_2 can be put in the form (cf. (5.8))

$$X_2=Xr_{k+1\,k+2}+Y$$

and

$$Y_2=X+Yr_{k+1\,k+2}$$

so that as $r_{k+1\,k+2}=-1$, $Y_2=-X_2$, and $\bar{Y}_2=-\bar{X}_2$. Since $r_{k\,k+1}=1$ (7.5) is then indeterminate. If the $k, k+1$ surface is the last surface, $X_2=1$ and $Y_2=0$, so that (7.5) reduces to (7.7), which is unity if $|r_{k\,k+1}|^2=1$, irrespective of whether $r_{k\,k+1}$ is complex or not. Therefore the generalization is in accordance with these energy considerations.

Vašiček[19] in his paper, by the assumption of "basic amplitudes," obtains the result for a two-film problem (Eq. (13) of Vašiček's paper).

$$r\exp i\delta=\left[\frac{r'+r''\exp -iX_1+r'''\exp -i(X_1+X_2)+r'r''r'''\exp -i(2X_1+X_2)}{1+r'r''\exp -iX_1+r'r'''\exp -i(X_1+X_2)+r''r'''\exp -i(2X_1+X_2)}\right], \quad (7.6)$$

where to convert to the notation of this paper

$$r\exp i\delta=\alpha_{14},\quad r'=r_{12},\quad r''=r_{23},\quad r'''=r_{34},\quad -X_1=2\phi_2,\quad -X_2=2\phi_3.$$

(7.6) differs from (4.3) in the phase of the terms in $r'r''r'''$ and $r''r'''$. If in (7.6) r''' is put equal to unity, and r', r'' and X_1, X_2 are considered to be real, multiplying (7.6) by its complex conjugate results in

$$|\ |^2=\left[\frac{\begin{array}{l}1+r'\{\exp i(X_1+X_2)+\exp -i(X_1+X_2)\}+r''\{\exp i(X_2)+\exp -i(X_2)\}+r'^2\\ \quad +r''^2+2r'r''\{\exp i(X_1)+\exp -i(X_1)\}+r'^2r''\{\exp i(2X_1+X_2)+\exp -i(2X_1+X_2)\}\\ \quad +r'r''^2\{\exp i(X_1+X_2)+\exp -i(X_1+X_2)\}+r'^2r''^2\end{array}}{\begin{array}{l}1+r'\{\exp i(X_1+X_2)+\exp -i(X_1+X_2)\}+r''\{\exp i(2X_1+X_2)+\exp -i(2X_1+X_2)\}+r'^2\\ \quad +r''^2+2r'r''\{\exp i(X_1)+\exp -i(X_1)\}+r'^2r''\{\exp i(X_2)+\exp -i(X_2)\}\\ \quad +r'r''^2\{\exp i(X_1+X_2)+\exp -i(X_1+X_2)\}+r'^2r''^2\end{array}}\right].$$

Since the coefficients of r'' in the numerator and denominator are different, as are also the coefficients of r'^2r'', r^2 is not unity for all values of r', r'', X_1, and X_2 when these are real and when $r'''=1$. A similar result is obtained when $r'''=-1$. So Vašiček's equation does not satisfy what appears to be a physical necessity, and it is suggested that his assumption of "basic amplitudes" is at fault. If the method of Vašiček is extended to three or more films, the number of terms in incorrect phase increases.

THE REVERSAL OF THE SYSTEM OF FILMS WHEN ALL THE MEDIA ARE NON-ABSORBING

In the case of an anti-reflection quarter wave film it is well known that the reflectivity of a glass surface coated with such a film is the same irrespective of whether the light is incident upon the film from the glass or from the air side. In the case of more complicated film systems the result is not so obvious, and to prove that the same result applies might not be so trivial as to be pointless.

Let all the media be non-absorbing so that all the rs and ϕs are real. Let there be n films and let the scheme corresponding to these films with air as the first medium be called scheme one. Multiply scheme one by $\exp -i(2\phi_2+\cdots+2\phi_{n+1})$, giving scheme two. Then the numerator drawn from scheme two is

$$N_{1\,n+2}\exp -i(2\phi_2+\cdots+2\phi_{n+1})$$

and the denominator is $D_{1\,n+2}$. Write down scheme two in the reverse order and let this give scheme three. Writing down a scheme in the reverse order does not affect the phase of the odd products but reverses the sign of the phase of the even products, so the numerator drawn from scheme three is $N_{1\,n+2}\exp -i(2\phi_2+\cdots+2\phi_{n+1})$ and the denominator is $\bar{D}_{1\,n+2}$. Multiply each element of scheme three by -1 and obtain scheme four. Since $r_{12}=-r_{21}$, this is equivalent to changing the order of the subscripts of each of the rs. Scheme four is the scheme which would be obtained if the system of films were reversed, i.e., the light incident on the opposite side, but with the sign of the phase of each element reversed. The numerator drawn from scheme four is $-N_{1\,n+2}\exp -i(2\phi_2+\cdots+2\phi_{n+1})$ and the denominator $\bar{D}_{1\,n+2}$. But the numerator is also $\bar{N}_{n+2\,1}$ and the denominator is also $\bar{D}_{n+2\,1}$. So

$$\bar{N}_{n+2\,1}/\bar{D}_{n+2\,1}=-N_{1\,n+2}\times\exp -i(2\phi_2+\cdots+2\phi_{n+1})/\bar{D}_{1\,n+2},$$

or

$$N_{n+2\,1}/D_{n+2\,1}=-\bar{N}_{1\,n+2}\times\exp i(2\phi_2+\cdots+2\phi_{n+1})/D_{1\,n+2}.$$

From this it can be seen that if all the media are non-absorbing, i.e., all the rs and ϕs real,

$$|\alpha_{1\,n+2}|^2=|\alpha_{n+2\,1}|^2.$$

ALL THE FILMS OF VANISHING THICKNESS

In view of the second of Lord Rayleigh's objections to the use of Stokes' treatment in the problem of the thin plate,[23] it is necessary to show that such a treatment does reduce to the correct coefficients when all the films are of vanishing thickness.

The Reflected Amplitude

If all the films are of vanishing thickness, irrespective of whether they are absorbing or not, $\exp i2\phi_k=1$, so that

$$N_{1\,n+2}=\tfrac{1}{2}[(1+r_{12})(1+r_{23})\cdots(1+r_{n+1\,n+2})-(1-r_{12})(1-r_{23})\cdots(1-r_{n+1\,n+2})]$$

and

$$D_{1\,n+2}=\tfrac{1}{2}[(1+r_{12})(1+r_{23})\cdots(1+r_{n+1\,n+2})+(1-r_{12})(1-r_{23})\cdots(1-r_{n+1\,n+2})].$$

If the electric vector is perpendicular to the plane of incidence, substituting from (2.3)

$$\alpha_{1\,n+2}=\frac{[(\mu_1c_1)(\mu_2c_2)\cdots(\mu_{n+1}c_{n+1})-(\mu_2c_2)(\mu_3c_3)\cdots(\mu_{n+2}c_{n+2})]}{[(\mu_1c_1)(\mu_2c_2)\cdots(\mu_{n+1}c_{n+1})+(\mu_2c_2)(\mu_3c_3)\cdots(\mu_{n+2}c_{n+2})]}.$$

Dividing numerator and denominator by $(\mu_2c_2)\times(\mu_3c_3)\cdots(\mu_{n+1}c_{n+1})$,

$$\alpha_{1\,n+2}=[\mu_1c_1-\mu_{n+2}c_{n+2}]/[\mu_1c_1+\mu_{n+2}c_{n+2}]=r_{1\,n+2}.$$

If the electric vector is in the plane of incidence, upon substituting from (2.5) and dividing the

[23] See reference 1, p. 67.

numerator and denominator by $(\mu_2 c_2)(\mu_3 c_3)\cdots \times(\mu_{n+1}c_{n+1})$,

$$\alpha_{1\,n+2}=[\mu_{n+2}c_1-\mu_1 c_{n+2}]/[\mu_{n+2}c_1+\mu_1 c_{n+2}]=r_{1\,n+2}.$$

The Transmitted Amplitude

If all the films are of vanishing thickness,

$$\beta_{1\,n+2}=t_{12}t_{23}\cdots t_{n+1\,n+2}/\tfrac{1}{2}[(1+r_{12})(1+r_{23})\cdots(1+r_{n+1\,n+2})+(1-r_{12})(1-r_{23})\cdots(1-r_{n+1\,n+2})].$$

If the electric vector is perpendicular to the plane of incidence, substituting from (2.3) and (2.4) it is easy to show that

$$\beta_{1\,n+2}=2\mu_1 c_1/[\mu_1 c_1+\mu_{n+2}c_{n+2}]=t_{1\,n+2},$$

and in the case of the electric vector in the plane of incidence by substituting from (2.5) and (2.6) the same result is obtained. The generalization therefore reduces to the correct values when all the films are of vanishing thickness.

ACKNOWLEDGMENT

I have to thank Professor A. M. Taylor for his constant encouragement and the readiness with which he has given me advice in the preparation of this paper.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 11 NOVEMBER, 1948

Coefficients of Reflection in the Ultraviolet of Magnesium Carbonate and Oxide

FRANK BENFORD, SALLY SCHWARZ, AND GWEN P. LLOYD
General Electric Research Laboratory, Schenectady, New York

(Received July 8, 1948)

The coefficients of reflection of magnesium carbonate and magnesium oxide have been determined for eight mercury lines in the ultraviolet. A spectrometer made up of a Rowland grating furnished a spectrum of light reflected from a part-sphere of the material under test. A photo-multiplier tube received the spectrum. The materials are both good reflectors, the oxide being especially good since it is over 0.90 at all points tested.

THE coefficients of reflection of magnesium carbonate and magnesium oxide have been determined for the eight mercury lines at 2400, 2536, 2804, 2967, 3023, 3126, 3341, and 3650 angstroms. The part-sphere method previously described[1] was used, and the carbonate blocks, after being tested as carbonate, were given 12 coats of oxide and then retested as magnesium oxide. The comparison was between 5/6 sphere and 4/6 sphere.

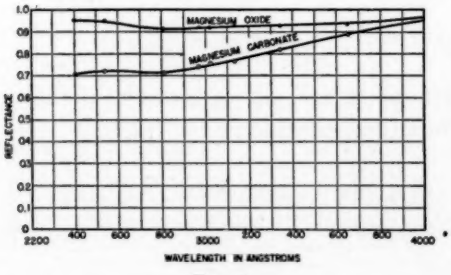


FIG. 1.

The source of energy was a Uviarc at 3.2 amperes and 145 volts. Radiation from the sphere entered a fixed slit spectrometer designed by Dr. Frank J. Studer. This instrument has a 15,000 line replica Rowland grating and a nine stage photo-multiplier tube as a receiving element. The voltage per stage was varied from 50 to 100 volts according to needs. The output of the tube was fed to a low sensitivity galvanometer.[2] This circuit has a high over-all sensitivity, but the stability is not perfect, and the testing procedure was almost identical with that employed when using a photometer. So far as possible, the readings were made in pairs; one with 5/6 sphere and then one with 4/6 sphere. In general, the

[1] Frank Benford, Gwen P. Lloyd, and Sally Schwarz, "Coefficients of reflection of magnesium oxide and magnesium carbonate," J. Opt. Soc. Am. **38**, 445 (1948).

[2] Frank J. Studer, "A method for measuring the spectral energy distribution of low brightness light source," J. Opt. Soc. Am. **37**, 288 (1947).

galvanometer readings, for the same conditions, would gradually increase throughout the day, with an occasional abrupt increase of five or ten percent. This made it necessary to arrange the work so as to eliminate the effects of time as far as possible.

The grating gives higher order spectra overlying the first-order spectrum that is used. To eliminate these higher orders, a piece of plate glass was used, as a filter, when working in the visible region. With this precaution, it was found that a reasonable reflection check could be gotten for 5461A, thus connecting the data of this test with the previous photometric test.

The carbonate coefficients decrease steadily from 0.890 at 3650A to 0.710 at 2400A, the shortest wave-length tested. The oxide (see Fig. 1) has coefficients of higher than 0.900 throughout the range of the test, with a low region near 2800A, and rising values at the shorter wave-lengths. In general, there is some similarity between the two materials, and both seem like usable reflectors in the region 2400A to 4000A.

July 1948 Meeting of the International Commission of Optics

THE International Commission of Optics of the International Union of Pure and Applied Physics held its first regular meeting at the Technical University in Delft, Holland, during the period July 12–17, 1948. There had been a preparatory meeting in Prague, Czechoslovakia, in June 1947, but without representation from the United States and some other interested countries. The I.C.O. is now formally affiliated with the I.U.P.A.P., and as a matter of fact this affiliation is particularly strong since Professor P. Fleury of Institut d'Optique, Paris, serves as the Secretary of both organizations.

Eleven countries were represented at the Delft meeting, the total number of official delegates being approximately thirty. The United States delegates were: Professor F. A. Jenkins of the University of California (Berkeley), Dr. D. B. Judd of the National Bureau of Standards, and Professor S. S. Ballard of Tufts College.

There were five very busy days of sessions. These were devoted largely to the presentation of reports by representatives of the various national committees, on topics assigned at the Prague meeting. Some of these were rather highly technical reports on a specialized field of optics, whereas others were progress reports or surveys of activities in a certain line, in the various countries: for example, the Swedish report on the manufacture of diffraction gratings and the general requirements for gratings throughout the world. Invited papers were presented by T. Smith of England, D. B. Judd of the United States, A. C. S. van Heel of Holland, and M. Françon of France. Interspersed throughout the sessions were discussions, sometimes of a very lively nature, of the general purposes and the proper activities and functions of the I.C.O.; in this respect the meeting was still of an organizational nature.

Visits were arranged to the local firms of Kipp and Zonen and the Oude Delft Optical Industry. There was also a visit to the Kamerlingh Onnes Laboratory of the University of Leiden. A most interesting all-day bus trip was taken to the Physical Laboratory of the University of Utrecht and the Research Laboratories of the Philips Lamp Works in Eindhoven.

At the final sessions, formal statutes were adopted for governing the I.C.O. These statutes were of course in accordance with applicable statutes of the I.U.P.A.P. Recommended functions for the I.C.O. were adopted for trial use and formal consideration at the next general meeting. These functions include the gathering and distribution of information on optical work and workers in the various countries and on the availability of optical materials and parts; an effort to agree upon a unified set of terms, symbols and definitions in optics; the preparation of suggestions for standards in optics, to be submitted to regularly constituted standardizing bodies; and the sponsoring of congresses on restricted fields of optics. It was emphasized that the I.C.O. should not act as a vehicle for publishing original research papers, but that full use should be made of existing scientific journals.

Representatives of the several national committees agreed to the preparation in their countries of reports on various technical topics, or accepted the responsibility for distributing questionnaires in certain specific fields. It was understood that the individual national committees would have maximum freedom in deciding when and how these items should be reported.

The officers elected at Prague to the Provisional Bureau were re-elected, and in addition two more Vice-Presidents were elected. The present Executive Committee is as follows:

President—T. Smith (England).
Vice-Presidents—J. Hrdlicka (Czechoslovakia), A. C. S. van Heel (Holland), S. S. Ballard (U.S.A.).
Secretary—P. Fleury (France).
Treasurer—A. Arnulf (France).

It is planned to hold the next regular meeting of the Commission in London in the summer of 1950.

STANLEY S. BALLARD,
Tufts College, Medford, Massachusetts

Light Scattering in Optical Glass*

D. H. RANK AND A. E. DOUGLAS**
Physics Department, Pennsylvania State College, State College, Pennsylvania

(Received August 9, 1948)

Raman spectra of two optical glasses (a borosilicate crown and a dense flint glass) have been obtained making use of unpolarized exciting light and also with parallel and perpendicular plane polarized exciting light. The chief characteristics of these spectra are that they consist essentially of continua with broad maxima in the continua. The main features of the Raman spectra of these glasses are strongly polarized. In the photograph made with parallel polarized exciting light, a very strong depolarized line only 30-cm^{-1} frequency shift from the exciting line makes its appearance in the spectrum of the flint glass.

An attempt has been made to find modified Brillouin components in the Rayleigh line. Photographs of λ-4047 and λ-4078 made in the third order of a 21-foot concave grating show no modified Brillouin components which are 1/100 as strong as the unmodified line.

The polarization of the Rayleigh scattering of these glasses has been investigated. The intensities of the polarized components were measured by means of photoelectric methods. Measurements made with polarized exciting light show definitely that ρ_h is somewhat less than unity, which definitely verifies the existence of a Krishnan effect in these glasses.

OPTICAL glasses exhibit many properties which at the present time are of a highly controversial nature. The purpose of this investigation is to examine some optical glasses making use of the various methods which light scattering provides for probing into the nature of materials.

The experiments performed fall into two general classes. (a) The nature and polarization of the Raman spectrum; (b) The nature and polarization of the Rayleigh scattering.

I. RAMAN SPECTRUM

The Raman spectrum of glasses has been investigated by numerous authors. A rather complete bibliography and discussion of the work previous to 1939 has been given by Hibben.[1] Kujumzelis[2] has made a rather extensive investigation of a number of glasses and has noted some of the unusual features of these spectra. More recently R. Norris[3] has made an extensive investigation of the Raman spectrum of a large number of Jena glasses. Norris has emphasized many of the unusual characteristics of these spectra.

In addition to the frequency characteristics of the Raman spectra of glasses, the polarization characteristics of these spectra should also be of importance to enable interpretation of the spectrum. We have not been able to find references in the literature to any work on the polarization of the Raman spectra of glasses in spite of the obvious importance of such measurements for interpretation of these spectra.

Part II of this paper, which is concerned with certain critical experiments on Rayleigh scattering, necessitated the use of glass samples which exhibit a high degree of optical perfection. We are greatly indebted to Dr. A. F. Turner of the Bausch and Lomb Optical Company for providing us with blocks of glass which had been specially selected for homogeneity, freedom from strain, and bubbles.

The two glasses investigated were DF-3, $N_D=1.621$, $V=36.2$ and BSC-2, $N_D=1.517$, $V=64.5$. These glasses were chosen because of their widely different optical and polarization properties with regard to Rayleigh scattering.

Raman spectra of the two glasses mentioned above were obtained photographically with a fast glass spectrograph which gives a linear dispersion of about 30 angstroms per mm at λ-4500A. The Hg arc light sources were of a low pressure type which give high intensity and extremely small continuous background.[4,5] Excitation of spectra

* This research was carried out on Contract N6onr-269, Task V of the Office of Naval Research.

** Now with National Research Council, Ottawa, Ontario, Canada.

[1] James H. Hibben, *The Raman Effect and its Chemical Applications* (Reinhold Publishing Corporation, New York), p. 404 (1939).

[2] T. G. Kujumzelis, Zeits. f. Physik **97**, 561, 100, 221 (1935).

[3] R. Norris, Proc. Ind. Acad. Sci. **14**, 178 (1941).

[4] D. H. Rank and J. S. McCartney, J. Opt. Soc. Am. **38**, 279 (1948).

[5] D. H. Rank, N. Sheppard, and G. J. Szasz, J. Chem. Phys. **16**, 698 (1948).

was by means of λ-4358, the short wave mercury radiation being removed by use of a sodium nitrite filter. Spectra were obtained with natural unpolarized exciting light and also with P and S polarized exciting light. The work with polarized exciting light essentially makes use of the chief features of the method of, "polarized incident light," described by us[6] previously.*** However, these weak spectra were recorded photographically instead of photoelectrically, and no quantitative depolarization measurements were made. Plates obtained were microphotometered with a Leeds and Northrup microphotometer. Smoothed tracings of some of the microphotometer curves are shown in Fig. 1. Tracings of plates obtained with natural unpolarized light are not included since they are qualitatively indistinguishable from those obtained with S polarized exciting light.

Reference to curves A and C, Fig. 1, shows features very similar to those shown by Norris[3] in his drawings of the intensity distribution in the spectra of the large number of glasses investigated by him. It must be pointed out that the chief features shown by the Raman spectra of glasses as exhibited by well exposed plates are continua with diffuse maxima in these continua. This feature of the spectrum is nicely demonstrated by Fig. 1C, where there is an almost complete break in the continuum at 850 cm^{-1}, which is comparatively sharp. (In comparing the curves of Norris with those of Fig. 1 it must be remembered that his curves are linear intensity plots and the microphotometer curves are logarithmic intensity plots. These logarithmic plots greatly reduce the heights of the intense peaks, especially since the microphotometer curves have been made from very intense plates in order to bring out the salient features of the spectrum.)

There seems to be agreement among numerous authors who have investigated glasses that one of the chief features of the Raman spectrum of glasses is a strong continuum extending from the exciting line to a long wave limit of from 400 to 500 cm^{-1} with a fairly sharp long wave edge. Norris[3] has shown by means of filters that this continuum is definitely a Raman effect and not fluorescence.

DF-3 Glass

This glass has a composition of about equal proportions by weight of SiO_2 and PbO, and about 10 percent of K_2O along with very small percentages of trace elements. The continuum which extends from the exciting line to a long wave limit at 520 cm^{-1} is extremely intense, appears to be of nearly uniform intensity throughout, and is devoid of structure. As can be seen from Fig. 1B, this continuum is very strongly

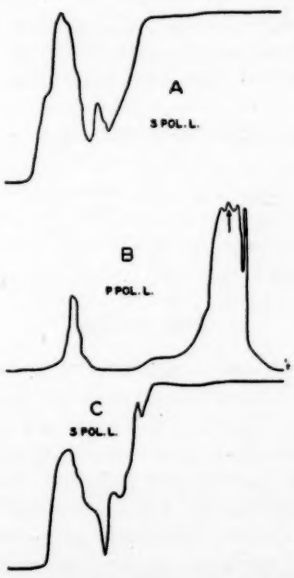


FIG. 1. Microphotometer curves of Raman spectrum of optical glass obtained by excitation with polarized exciting light. (A) Raman spectrum of DF-3 glass excited by ⊥-polarized exciting light; (B) identical to A excited by ||-polarized light; (C) Raman spectrum of BSC-2 glass excited with ⊥-polarized exciting light. The arrow in curve B denotes the exciting line Hg 4358 angstroms.

[6] A. E. Douglas and D. H. Rank, J. Opt. Soc. Am. **38**, 281 (1948).

*** The chief relevant features of the method of polarized incident light are: (a) The method is essentially free of convergence errors; (b) transmission of the spectrograph for the two kinds of plane polarized light does not enter into the results since the light being analyzed by the spectrograph has the same identical polarization characteristics for the two exposures; (c) the method measures ρ_S, which is connected with the more usually determined ρ_n by the purely geometrical relationship $\rho_n = 2\rho_S/1+\rho_S$.

Essentially highly polarized features of the spectrum will be practically extinguished when P polarized exciting light is used, while highly depolarized features will be enormously enhanced in photographs obtained with this type of exciting light. This last characteristic of the method is of particular importance for observing detail in the spectrum in the neighborhood of the very highly polarized Rayleigh line and its associated continuum.

TABLE I. Frequency shifts and polarization characteristics of the Raman spectrum of two glasses. V. S.—very strong, Dep.—depolarized, Pol.—polarized, S.—strong, W.—weak, Lim.—limit, M.—medium, Sh.—sharp.

Glass DF-3	Glass BSC-2
30 cm^{-1} V. S. Dep.	550 cm^{-1} Lim. cont. Pol.
520 cm^{-1} Lim. cont. V. S. Pol.	550–820 W. cont. Pol.
790 cm^{-1} M. Dep.	630 cm^{-1} M. Sh. Pol.
870–950 W. cont. Dep.	875–1200 M. cont. Pol.
980 cm^{-1} S. Max. Dep.	Max. 1050 cm^{-1}
1050 cm^{-1} S. Max. Pol.	
1070–1180 W. cont. Dep.	

polarized. The strong band in the 1000-cm^{-1} region is definitely a double band with maxima at 980 cm^{-1} and 1050 cm^{-1}. The spectrum obtained with P polarized exciting light shows the 980-cm^{-1} band to be practically completely depolarized, while the 1050-cm^{-1} band is completely polarized. Probably the most interesting feature in the spectrum of this glass is the completely depolarized frequency which shows up at 30-cm^{-1} shift from the exciting line. This frequency can clearly be seen in Fig. 1B to the left of the exciting line, which is indicated by an arrow. Although the 30-cm^{-1} frequency is very intense it is only possible to observe it by virtue of the high degree of polarization of the Rayleigh scattering. At the present time any explanation of the appearance of this low lying frequency would be pure speculation. This frequency arises obviously from a type of rotational oscillation of groups of atoms in the glass, but the nature or arrangement of atoms or groups responsible for its appearance would have to be decided by further experiments with glasses of different compositions.

BSC-2 Glass

This glass has an approximate composition of 70 percent SiO_2 and about 10 percent each of K_2O, Na_2O, and B_2O_3. The main features of the spectrum of this glass are shown in Fig. 1C. A photograph obtained using P polarized exciting light showed that the Raman spectrum of this glass was more or less completely polarized since no maxima of any kind appeared in the P polarized photograph. In addition, the low lying frequency observed in the lead glass when the spectrum was excited by P polarized light does not appear for the borosilicate glass. Frequency and polarization characteristics for the two glasses investigated are given in Table I. From the behavior of frequency and polarization characteristics of the Raman spectrum of these glasses it might be concluded that investigation of other glasses of systematically varied compositions might lead to fruitful results.

II. RAYLEIGH SCATTERING

The fine structure of Rayleigh scattering has been studied by numerous investigators in liquids and crystals. The observed frequency separation of the Doppler components is known to obey the classical formula

$$\Delta\nu = \pm 2\nu(v/c)u \sin(\theta/2)\cdots, \qquad (1)$$

where ν is the frequency of the exciting light, v the velocity of sound in the liquid, c the velocity of light, u the refractive index, and θ the angle of scattering. Landau and Placzek[7] have given a formula which gives a relationship between the intensity of modified and unmodified Brillouin components which might be expected. More recently Gross[8] has given a derivation of the Landau Placzek formula,

$$I_c/2I_B = \gamma - 1\cdots, \qquad (2)$$

where I_c is the intensity of the unmodified component and I_B is the intensity of the modified or Doppler component. γ is the ratio of the specific heat at constant pressure to the specific heat at constant volume. One of us[9] has shown in the case of ethyl alcohol and acetone (two liquids which have a low viscosity and also a very weak rotational Raman effect in the liquid state) that the Landau Placzek formula adequately expresses the experimental observations. From Eq. (2) it would be expected that the unmodified component should completely disappear from the fine structure pattern of the Rayleigh line produced by a solid. It is well known from the work of numerous investigators that the unmodified line is absent in the Rayleigh line produced by crystals. Gross (1930) and Ramm (1934), making use of interferometers, have reported unsuccessful attempts to find modified Brillouin components in the Rayleigh scattering of glass.

[7] L. Landau and G. Placzek, Physik. Zeits. Sowjetunion **5**, 172 (1934).
[8] E. Gross, Acta Physicochimica U.R.R.S. **20**, 459 (1945).
[9] D. H. Rank, J. S. McCartney, and G. J. Szasz, J. Opt. Soc. Am. **38**, 287 (1948).

Venkateswaran[10] has shown that the intensity of the modified Brillouin components compared to the unmodified component decreases rapidly with increasing viscosity in liquid phenol. In the light of Venkateswaran's experiments, since glasses are extremely viscous liquids, it might be expected that the modified Brillouin components are completely absent in the Rayleigh line.

Since the velocity of sound is large in a glass the Doppler displacements will be of the order of 0.14A for $\lambda = 4000$A in contradistinction to 0.035A for most of the common liquids. This situation allows the use of the diffraction grating for observation of possible fine structure of Rayleigh scattering. The use of the grating permits the observation of much weaker satellites than is possible with the interferometer.

We have obtained photographs of the Rayleigh scattering from both DF-3 and BSC-2, making use of the third order of a 21-foot grating. Photographs were obtained for both λ-4047 Hg and λ-4078. The photographs were obtained making use of a cylindrical lens in front of the plate to remove the astigmatism of the grating in the manner described by one of us[11] previously. The photographs show that no modified Brillouin components as much as 1/100 the intensity of the unmodified component are present in the Rayleigh scattering of these glasses. These results tend to add additional evidence to the supposition that in effect glasses behave more like supercooled liquids than rigid solids in spite of their elastic properties.

Polarization of Rayleigh Scattering

Krishnan[12] has discovered an effect bearing his name over which considerable controversy has arisen. The chief feature of the phenomenon is that in the Rayleigh scattering of certain liquid mixtures and solids, light polarized with the electric field vibrating in the plane parallel to the plane containing the incident and scattered beams is not completely depolarized as would be expected in the case of true molecular scattering. This phenomenon was independently discovered by Rousset[13–15] but was explained by him as being essentially due to secondary scattering since the ρ obtained varied with the geometrical shape and size of the volume of the liquid illuminated. Krishnan has made extensive investigations on this phenomenon in opalescent liquid mixtures and also in glasses.[16,17]

This "anomalous" polarization has been attributed by Krishnan to the presence of molecular clusters not small compared to the wave-length of light. Krishnan[18] has derived a relationship between the depolarization factors for unpolarized light, ρ_u, and the two kinds of plane polarized light, ρ_v and ρ_h. This relationship is derived from the Theory of Reciprocity of Lord Rayleigh.[19] Mueller[20] has given a thorough discussion of the Krishnan effect and has shown that the Reciprocity Theorem is always valid. Mueller has developed a theory of the Krishnan effect in glass and has obtained good agreement with Krishnan's experimental observations. The expression giving the relationship of the three ρ's according to Krishnan is

$$\rho_u = (1+1/\rho_h)(1+1/\rho_v). \qquad (3)$$

The criticism of Krishnan's work has been of a dual nature. The Krishnan effect occurs, in general, only in cases where the scattered light is

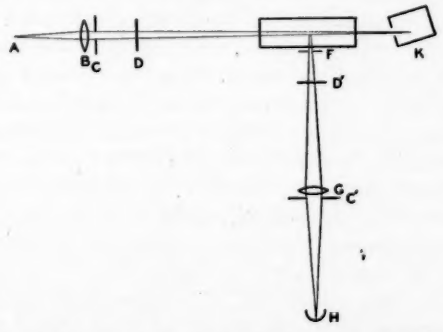


FIG. 2. Schematic diagram of apparatus used to measure depolarization factors of optical glasses. *A*-automobile headlight bulb, *B* and *G*-condensing lenses, *C*, *C'*, and *F*-aperture stops, *K*-light trap, *D* and *D'*-polaroids mounted on divided circles, *H*-multiplier photo-tube.

[10] C. S. Venkateswaran, Proc. Ind. Acad. Sci. **15**, 322 (1942).
[11] D. H. Rank and J. A. Van Horn, J. Opt. Soc. Am. **36**, 325 (1946).
[12] R. S. Krishnan, Proc. Ind. Acad. Sci. **1**, 211 (1934).
[13] A. Rousset, Comptes rendus **198**, 2152 (1934).
[14] A. Rousset, Comptes rendus **199**, 716 (1934).
[15] A. Rousset, thesis (Paris, 1935).
[16] R. S. Krishnan, Proc. Ind. Acad. Sci. **1**, 915 (1935).
[17] R. S. Krishnan, Proc. Ind. Acad. Sci. **3**, 211 (1936).
[18] R. S. Krishnan, Proc. Ind. Acad. Sci. **6**, 21 (1938).
[19] Lord Rayleigh *Theory of Sound* (1926), Vol. 1, p. 93.
[20] H. Mueller, Proc. Roy. Soc. **166A**, 425 (1938).

TABLE II. Depolarization factors for optical glasses; u-refractive index.

u	ρ_v%	ρ_u%	ρ_h% Obs.	ρ_h% Calc.	Density
1.621	1.1	2.5	90	77	3.67
1.517	6.1	14	85	70	2.53
*1.4925	12	24	82	81	2.3
*1.5269	3.6	8.1	80	75	2.5
*1.6473	2.3	5.2	75	76	3.9
*1.6163	1.6	3.7	80	74	3.5

* Measurements of Krishnan (see reference 17).

highly polarized. This might lead to anomalous values of ρ_h as a result of convergence errors. From examination of the experimental details of his work there seems to be no doubt that convergence error plays no part in the results obtained. This is especially true in the work on glasses. The second objection, secondary scattering (i.e., the explanation of the anomalous value of ρ_h originally given by Rousset) for the critical liquid mixtures is fundamentally much harder to disprove.[21] However, even if the secondary scattering explanation would have some justification in the critical liquid mixtures it could not possibly explain the phenomenon in the case of optical glass where the primary scattering is many times weaker, thus reducing secondary scattering to complete insignificance.

Krishnan's experimental method for determining depolarization is essentially the method of Cornu and is, of course, not completely objective. The method for determining ρ_h involves the matching of two extremely weak fields of nearly equal intensity by rotation of a Nicol. Krishnan[22] has made use of a more objective photometric method involving photographic photometry to show the "anomalous" value of ρ_h for several optical glasses. This same method gave a normal result, namely, $\rho_h=1$ for liquid ether.

We have made use of modern photoelectric measuring instruments such as used by us[6] recently to attempt to obtain precise and objective values for ρ_h for the two glasses studied in this paper. We have also measured ρ_u and ρ_v by this same objective method. The photoelectric method will give precise values of ρ_h, since the two polarized components to be measured have nearly equal magnitudes. However, it seems that the Cornu method will give better precision in the determination of ρ_u and ρ_v, since in this case the magnitude of the strong component is very large compared to the weak component.

A diagram of the apparatus used by us is shown in Fig. 2. The effective aperture of the lenses used was about $F=50$ in the glass. At this aperture ratio the convergence error correction is completely negligible. It can be seen that these large blocks of glass provide ideal optical crosses. Extreme precautions were taken to provide baffles and blackening of the glass at all places where parasitic light might originate to vitiate the measurements. The measurements on the flint glass were made with essentially monochromatic light provided by means of a "green line" interference filter and also by means of white light. Both sets of measurements gave essentially the same answers. The measurements on the crown glass were made only with white light because of the much lower intensity of the scattered light. Results obtained are given in Table II, along with results obtained by Krishnan for glasses of similar nature.

The results we have obtained given in Table II do not give a good check between ρ_h obs. and ρ_h calc. However, if one examines Eq. (3) it can be seen that ρ_h calc. is very sensitive to the accuracy of determination of ρ_u and ρ_v. As an example, if ρ_v for the first glass given in the table were 1.2 instead of 1.1 the calculated value of ρ_h would be 90 instead of 77. Our measurements of ρ_v and ρ_u can be expected to have a very small error in the last place, which can easily account for the discrepancy between calculated and observed values of ρ_h. Likewise we feel certain that our measurements of ρ_h have a very small error in the last place and believe they indicate definitely the ρ_h is less than unity for the optical glasses investigated, which is both qualitatively and quantitatively in agreement with Krishnan's measurements.

[21] B. Mookerjee, Ind. J. Phys. **12**, 15 (1938).
[22] R. S. Krishnan, Proc. Ind. Acad. Sci. **8**, 442 (1938).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 11 NOVEMBER, 1948

An Illuminant-Stable Color Vision Test

II. Experimental-Statistical Evaluation

ELLIS FREEMAN AND M. A. ZACCARIA
The University of Texas, Austin, Texas

(Received July 1, 1948)

Previously the physical specifications and the psychophysical requirements were presented for six inks from which an illuminant-stable red-green color vision test could be constructed. The test of 17 polychromatic plates was prepared and administered to 100 color blind and 100 normal subjects under illuminants ranging from minus-blue to 14000°K. The present paper reports the results of the evaluation of the test in actual practice. Theoretical expectations have been realized in that the test is illuminant-stable, discriminatory between normals and deficients, properly varied in relative difficulty of individual plates for normals and for color blinds, yielding scores well distributed in terms of the probability function, and with a minimum of overlapping between normals and color blinds.

I. INTRODUCTION

PAPER I[1] of this project gave the specifications for a printed polychromatic red-green test which would maintain stability, as a condition of diagnostic integrity, under the widest practically conceivable range of color temperatures of illumination. It also established the colorimetric specifications of six inks capable of meeting all the psychophysical requirements of such a test. The present paper reports an evaluation of the test in practice, as administered to 100 normal and 100 color blind subjects. For those who may be puzzled by this rich supply of color blind material, it may be stated that they were volunteers who responded to an advertisement for color blind subjects in the university paper. No slow and arduous screening was required, since great numbers of male veterans now on the campus had been tested in the armed forces. Thus each volunteer knew his qualifications in advance. On the conservative basis of four percent incidence among males, a pool of 500 color blinds was expected, of which the first 100 to appear were used.

The importance of employing the standard illuminant for which the customary polychromatic printed test is designed is almost universally ignored in practice. Although this indifference and the consequent failure to detect 50 percent of the deficients by the criteria of the tests themselves[2] has been recognized, the responsibility of non-standard illumination as a source of error may be more fully appreciated from Fig. 1. This figure represents color temperatures recorded at somewhat irregular intervals over a period of forty-eight hours in Army and Navy recruiting quarters in Austin, Texas, while they were open for business. The range is from 2600°K to 10,500°K, with non-standard frequencies far outnumbering those even approximating 6500°K (Illuminant *C*), the standard temperature for which their test has been designed. Moreover, Fig. 1 sharply demonstrates that a test designed for any illuminant other than *C* as standard would enjoy no useful probability of avoiding error without control of lighting. But since lighting cannot in actual practice be controlled, the utility of an illuminant-stable test is apparent.

II. APPARATUS

In order to evaluate the present Illuminant-Stable test (hereafter referred to as the I-S test), a range of illuminants from minus-blue to 14,000°K was set up in a completely dark room. More specifically, these illuminants were:

Illuminant *I*, minus-blue (Wratten filter No. 16 with 75-w lamp)
Illuminant *II*, 2000°K (500-w lamp operated at 60 volts)

[1] Ellis Freeman, "An illuminant-stable color vision test. I," J. Opt. Soc. Am. **38**, 532 (1948).

[2] D. Farnsworth and J. D. Reed, *A Survey of Methods Used in Administering Pseudo-Isochromatic Test Plates for Color Vision*, Color Vision Report No. 3 (Bu. Med. x-260 (Av-148-C), Nov. 6, 1943), p. 1.

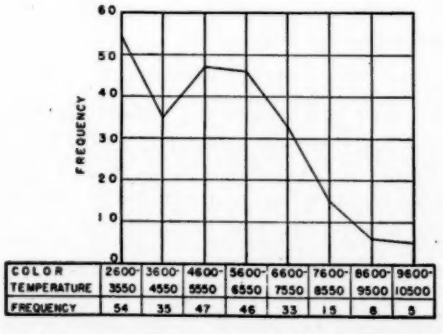
Illuminant *III*, 2850°K (Illuminant *A*, 75-w lamp)
Illuminant *IV*, 6500°K (Illuminant *C*, Macbeth filter with 500-w lamp)
Illuminant *V*, 14,000°K (special Macbeth filter with 3 "daylight" lamps of 200 w each).

Each illuminant was adjusted, by varying its vertical distance from the target, to deliver 25 footcandles onto the surface of the test plate. The stand for the test plates was so constructed as to admit one plate at a time for a 2-second presentation in random order, normal to the line of regard, 30 inches from the subject, and 45° to the illumination.

The American Optical test (hereafter referred to as the AO test) used for comparison with the I-S test consisted of the 18 plates selected as an improved and abridged edition of the original by Hardy, Rand, and Rittler.[3]

The I-S test consisted of 17 plates, embodying in various combinations the six inks *a*, *b*, *c*, *d*, *e*, *f* described in Paper I. After preliminary exploration it was found that this number of plates allowed at once for the maximum efficiency of administration coupled with an adequate sampling of triple, double, and single combinations of numerals on the plates. It also allowed adequate representation of each of the six inks in the series of plates.

An unusual and extremely desirable feature of the I-S test is that all 17 plates are identical in pattern. This uniform pattern is such that it allows from one to three numerals to be represented on each plate with no variation other than that of colors. By means of this uniform pattern all extraneous cues which may assist cheating are avoided.



COLOR TEMPERATURE	2600-3550	3600-4550	4600-5550	5600-6550	6600-7550	7600-8550	8600-9500	9600-10500
FREQUENCY	54	35	47	46	33	15	8	8

FIG. 1. Frequency distribution of color temperatures in actual and typical testing rooms of the armed forces.

[3] L. H. Hardy, G. Rand, and M. Rittler, "A screening test for defective red-green vision," J. Opt. Soc. Am. **36**, 610 (1946).

III. PROBLEM AND PROCEDURE

Since the over-all purpose of this study is an evaluation of the I-S test, the problem breaks down to: (a) determining its stability, as a condition of diagnostic integrity, under a sufficiently wide range of illuminants, (b) its validity in discriminating between the color blind and the normal, and (c) the relative difficulty of the plates. The third item suggested also that since a completely new kind of test was being devised, an effort should be made to introduce a range of difficulty among the plates for the normals as well as for the color blinds, in order to approximate the probability function and in general meet the criteria of sound test construction. This characteristic, absent from the conventional tests in common use in which normals cluster around perfect scores, is attainable without, as becomes apparent with the I-S test, sacrificing its power of discriminating between, or dichotomizing, the normals and the color blinds.

Five groups of red-green color blinds and five groups of normals were used in this study. Each group consisted of 20 subjects—in all 100 color blinds and 100 normals. The former were all male, while nearly two-thirds of the latter were female. The normals were so classified in accordance with the criterion proposed by Hardy, Rand, and Rittler.[3]

The subject was given the AO and the I-S test under each of the five illuminants, the latter being varied in order, as described below, for the purpose of equalizing the effect of learning. In conclusion each subject was retested with each test under the illuminant first used. Subjects in group *IC* (color blinds in group *I*) and in group *IN* (normals in group *I*) took the tests under illuminants *I*, *IV*, *III*, *II*, *V*, successively, and the retest under *I*; groups *IIC* and *IIN*, under *II*, *III*, *IV*, *V*, *I*; groups *IIIC* and *IIIN*, under *III*, *II*, *V*, *I*, *IV*; groups *IVC* and *IVN*, under *IV*, *V*, *I*, *II*, *III*; groups *VC* and *VN*, under *V*, *I*, *II*, *IV*, *III*. Testing was done under these varied sequences of illumination in order to

TABLE I. Frequency distributions of maximum differences of error scores under Illuminants *II**, *III*, *IV*, *V** for color blinds (*C*) and for normals (*N*) on the AO and the I-S tests.

	Frequency			
Difference	AO *C*	AO *N*	I-S *C*	I-S *N*
11	1			
10				
9	4			
8	8			1
7	11			
6	17			2
5	14		2	6
4	8		6	12
3	14	3	17	25
2	12	7	18	31
1	11	41	26	22
0		49	31	1
No. of subjects	100	100	100	100
Mean	4.7	0.6	1.5	2.7
S.D.	2.4	0.7	1.3	1.4

TABLE II. Frequency distribution of test-retest differences of error scores under Illuminants *I*, *II*, *III*, *IV*, *V* for color blinds (*C*) and for normals (*N*) on the AO and I-S tests.

	Frequency			
Difference	AO *C*	AO *N*	I-S *C*	I-S *N*
7				3
6			1	3
5	1		1	6
4	7		1	11
3	7	2	1	17
2	15	4	12	17
1	21	15	19	24
0	26	77	53	15
−1	16	1	6	2
−2	3	1	6	2
−3	4			
No. of subjects	100	100	100	100
Mean	0.7	0.3	0.8	2.2
S.D.	1.7	0.7	1.4	1.9

expose each plate of the two tests to the same total practice effect arising from repetition as such and possibly from differences resulting from the order of succession of illuminants.

For scoring, an incorrect reading of one or more digits of a figure on a plate was counted a single error. To be counted right, however, a response had to be correct for all digits on the plate. In the case of the AO test, if two different responses were given by an individual for one plate, as happened because of the ambiguity of some of the German-shaped numerals or on account of the intentional equivocality of others which could confuse normals as well as the deficients whom they were designed to deceive, the subject was required to declare the number "clearest" to him and scored accordingly.

Since one primary objective was to determine the stability of the I-S test under the selected range of illuminants, tabulations were made for each color blind and each normal of errors on each plate under each illuminant. To meet the criterion of stability an I-S plate would, ideally, have to elicit a consistent response from a given subject under all illuminants. This means that were a given plate failed by a given subject under one illuminant, he would fail it under all, and if passed anywhere, it would be passed everywhere. Therefore, if the results for all subjects were massed for a given plate, the total number of errors under each illuminant should ideally be constant. If, as was bound to be the case, some deviations did occur in practice, a measure of illuminant stability, applicable to each plate, had to be devised. After preliminary investigation a difference range of no more than 15 for color blind and normal was selected as a suitable criterion of illuminant stability, under all five original illuminants, from minus-blue through 14,000°K. Therefore the worst permissible performance of any one plate was a difference of 15 between the maximum of errors under one illuminant and the minimum of errors under another for all 200 subjects. Since each subject read each plate five times (once per illuminant), the total of responses is 1000, against which this maximum permissible deviation of 15 is to be considered.

However, it is even less, and, therefore, better, namely, 8, for the more practically representative gamut of the four illuminants (*II**, *III*, *IV*, *V**) discussed hereafter and mentioned in the legend of Table I.

Both the I-S and AO test were subjected to like administration and statistical treatment, so that direct comparison could be made between the one which is designed for illuminant stability and the other which is not. Comparison was also made as to test-retest reliability of the two.

The diagnostic power of each plate of the I-S test was another important consideration. In this respect, a useful index of validity for a test item is some type of coefficient of correlation. Here the Pearsonian *r* was used.

Next, each plate of the I-S test was ranked according to the order of difficulty. In this connection difficulty for normals as well as for color blinds was taken into account.

Finally, comparisons were made between the I-S and AO tests in respect to the distribution and overlapping of scores of color blinds and of normals. This indispensable consideration of sound test construction has hitherto been ignored in the *construction* of the conventional tests now in use. In subsequent evaluations, as by Hardy, Rand, and Rittler,[3] the seriousness of the oversight has become apparent.

IV. RESULTS

According to Fig. 1, the range of color temperatures in typical testing rooms was from 2600°K to 10,500°K, a span of 7900°K. Illuminants II to V, running from 2000°K to 14,000°K, cover a range of 12,000°K, far in excess of what has been found in testing rooms which represent ordinary conditions. It was, for this and other reasons which will appear, held desirable to interpolate scores for two assumed illuminants for each subject in order to get results conformable with two more moderate and practical extremes, so that the behavior of the tests under ordinary illumination could be determined. Such an interpolation is based on the assumption that test scores (errors) would run in a continuum between Illuminants *II* and *V*. Moreover, this restricted range of illumination places much less severe demands upon the AO test, since the more it is exempted from extremes of illumination the more valid it remains. At the same time it deprives the I-S test of the opportunity of exhibiting its validity at the far extremes to which its stability extends, as shown in Table III and in other extensive experimental data of which this paper is a condensation.

The two new assumed extremes selected were 2425°K and 10,250°K (hereafter referred to as Illuminants *II** and *V**, respectively). The first falls approximately midway between Illuminants *II* and *III*, and the other between *IV* and *V*. Together they establish a range of 7825°K, and, with *III* and *IV* added, satisfactorily duplicate conditions recorded at the recruiting stations, as it was at this juncture intended to do. The probable score for each subject under each of the two assumed illuminants was then obtained by interpolation.

Table I shows the frequency distributions of maximum differences of error scores for color blinds and for normals on both tests, under Illuminants *II**, *III*, *IV*, *V**. The table shows that: (a) the I-S test is consistent in means and standard deviations for color blinds and normals, as the AO test is not; (b) the means and standard deviations for the I-S test are smaller; (c) only three percent of the normals and none of the color blind have score deviations on the I-S test of more than 5; whereas none of the normals and forty-one percent of the color blind have score deviations of more than 5 on the AO test. These facts indicate that the I-S test is illuminant-stable. They also confirm the already established fact of the instability of the AO test, which, designed for a standard illuminant, could understandably not behave otherwise.

Table II shows the frequency distributions of the test-retest differences for the color blinds and the normals on both tests, together with their means and standard deviations. These should give a measure of practice effect. Very little of such effect could be noted in either group. The

TABLE III. For the I-S test, frequency of errors for color blinds (*C*) and for normals (*N*); the discriminatory power of each plate in terms of *r* (an *r* of 0.18 being significant to the 1 percent level of confidence); and the difficulty rank of each plate (1 being the easiest and 17 the hardest). For the frequency of errors the *lowest number for the color blinds* and the *highest for the normals* under any one of the five illuminants actually employed was selected to represent each plate. Thus the I-S test is here subjected to the most rigorous and unfavorable conditions.

	% of subjects erring on plate			
Plate	*C*	*N*	*r*	Difficulty rank
1	75	4	0.73	1
2	95	78	0.35	17
3	86	7	0.77	3
4	79	41	0.39	10
5	75	12	0.63	2
6	99	44	0.75	14
7	86	15	0.69	7
8	88	11	0.74	5
9	98	16	0.83	8
10	96	33	0.70	12
11	84	16	0.67	6
12	96	63	0.52	16
13	98	57	0.63	15
14	97	20	0.79	9
15	94	3	0.87	4
16	94	38	0.64	13
17	99	22	0.83	11

I-S normals better their retest scores on the average by only 2.2 points, while all the others change even less. From this it may be concluded that both the I-S and the AO tests have high reliabilities.

Table III gives the frequency of incorrect responses of the 100 color blinds and of the 100 normals, the discriminatory power, and the difficulty rank for each plate of the I-S test. To subject this test to the severest conditions, the lowest number of incorrect responses for color blinds and the highest number for normals for each plate under any of the five illuminants actually used were taken as the basis of computation. An r of 0.18 is significant to the 1 percent level. It will be noted that the lowest r is 0.35 and the highest 0.87. Over one-half the plates have an r of 0.70 or higher. Further significance is lent by the fact that this r approximates a biserial correlation coefficient. From these circumstances it may be concluded that the 17 plates of the I-S test as a whole, and each plate separately, discriminates very significantly between color blinds and normals.

In Fig. 2 are recorded and graphed the frequency distributions of error scores on the I-S and AO tests for color blinds and for normals under Illuminants *II**, *III*, *IV*, *V**. The mean and the standard deviation for each group under each test are recorded below the chart. The data are self-explanatory.

In considering the nature of the curves, it should be borne in mind that our subjects were not an unselected characteristic sampling of the population, since our color blinds were 50 percent of the total, and badly defective at that. From an unselected population one could expect some resemblance to a normal distribution, in which the preponderance of normals would show a wider range of scores than do the color blinds on any test. The present scores range from 0 through 4 (Hardy, Rand, and Rittler criterion) for the normals on the AO test, with 82 percent scoring 0 errors. The same individuals range from 0 through 12 on the I-S test. The color blinds score from 0 through 18 (appearing as normals under non-standard illumination) on the AO test. If 11 were set as the critical score for the color blinds on the I-S test, three normals

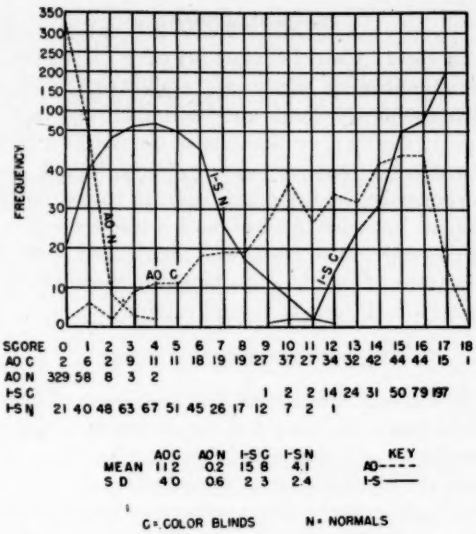


FIG. 2. Frequency distribution of error scores for color blinds (*C*) and for normals (*N*) on abridged (18 plate) AO pseudoisochromatic test and on illuminant-stable (I-S) test.

(hitherto so classified on the basis of the AO test) would have to be classified as color blind, and three color blinds as normal. With the AO test and five errors as the critical score, thirty color blinds would have to be classified as normal.

With a characteristic unselected sampling of the population the frequency of color blinds on the I-S and AO tests would be less than 5 percent of what is shown on the curves. If a correction were made here for the 50-50 sampling used, it would be seen that the I-S test scores would become more normally distributed, while the AO test scores would nearly all be bunched together at 0. This fact suggests that the AO test is much too easy for the normals to distribute themselves suitably in accordance with the probability function.

It may also be properly assumed that the means for the color defectives on both tests would have been lowered had a random selection of subjects been used. In that case the results would have been much more unfavorable to the AO test than they are here. There would be considerably more color defectives making low error scores, whereas on the I-S test they would distribute themselves close to the critical score of 11.

V. CONCLUSIONS

In the 17 plate I-S test: (a) each plate meets the condition of illuminant stability; (b) each plate has a very significant discriminatory power; and (c) the relative difficulty of the plates is well distributed, from very easy to fairly difficult for the normals, and from fairly difficult to extremely difficult for color blinds. The I-S test scores are, furthermore, as a whole well distributed, with a minimum of overlapping between those of the normals and those of the color blinds.

VI. SUPPLEMENTARY NOTE

In footnote 4 of Paper I[1] it was observed that ink *a* under Illuminant *Z*, as shown in Fig. 6A for the Protanopic only, just barely departed from the confusion zone of the other five inks. The explanation of the expedient for overcoming this minor difficulty, reserved for the present Paper II, now follows. First, the insignificance of the delinquency of this ink is obvious from the fact that none of the color temperatures recorded at the typical recruiting testing sites, as shown in Fig. 1 of the present paper, ever rose above 10,500°K, whereas Illuminant *Z* is 13,700°K. However, as a precaution against the remote possibility of an illuminant as high as *Z*, ink *a* was never used for a monochromatic ground or for the figure of a single digit plate. When used as ground, it was alternated with one or two other inks. When used as digit, it was accompanied by one or two other digits of different inks, selected from the remaining five inks which behaved properly under Illuminant *Z*. Since the subject was thus always faced by other inks which he was compelled to discriminate in order to arrive at a correct response, it is manifest that even under Illuminant *Z* the discernment of a portion of a figure through the assistance of ink *a* would not yield a correct response.

Similarly, it was observed in footnote 5 of Paper I that under Illuminant *A* ink *f*, and under Illuminant *Z* ink *a*, showed a slight excess over the permissible brightness difference between inks. The expedient for resolving this minor difficulty is identical with the above. Indeed, the difficulty here with ink *a* under Illuminant *Z* was overcome simultaneously with the procedure above for neutralizing the effect of its trespass into the adjacent confusion zone. The like brightness excess of ink *f* under Illuminant *A* was similarly resolved by alternating it with one or two other inks for ground, and by accompanying it by one or two other digits of other inks for figure. Thus, in general terms, the matter resolved itself quite simply into the recognition of a certain loss of degrees of freedom in making combinations of inks, so that some deliberation, instead of random combination of inks, was necessary.

Short Wave-Length Cut-Off Filters for the Ultraviolet*

ARNOLD M. BASS
Department of Physics, Duke University, Durham, North Carolina

(Received June 2, 1948)

A number of organic compounds have been investigated for possible use as short wave-length cut-off filters in connection with fluorescence studies in the near ultraviolet region (3000–2300A). The absorption of these filter compounds has been studied, in the pure liquid state and in isooctane solutions of various concentrations, by means of a Beckman DU spectrophotometer. It was found that for a number of substances the cut-off wave-length can be shifted over a fairly wide range by varying the concentration of the solution. Curves are given for phenol, benzene, cumene, and thiophene. A table lists the cut-off wave-lengths for these compounds and also for chlorobenzene, dihydropyran, and mesitylene.

IN the course of some studies on fluorescence spectra in the near ultraviolet region (3000–2300A) a need arose for filters which would transmit only a narrow band of wave-lengths, insuring a high degree of monochromasy in the exciting light. Since such filters are not available as yet for the region under investigation, it was decided to use cut-off filters which would transmit only those wave-lengths longer than a given value. In addition, it was desired to be able to vary the position of the cut-off wave-length. A number of filters are known for use in the desired region,[1–3] but little information was found on the variation of the transmission characteristics of these filters with changes of the various conditions such as concentrations, filter thicknesses, etc. It is known[4] that many organic compounds,

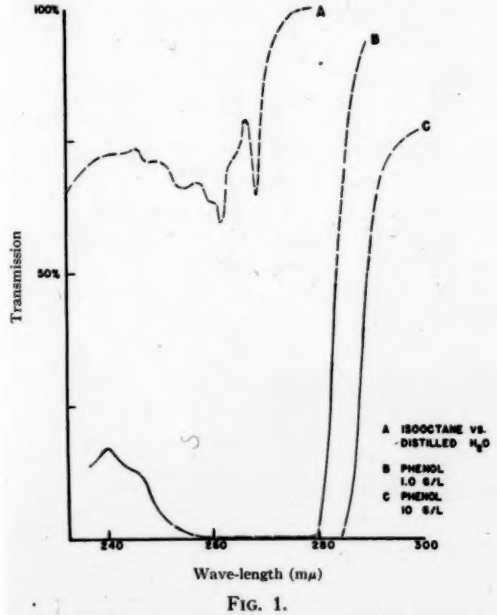


FIG. 1.

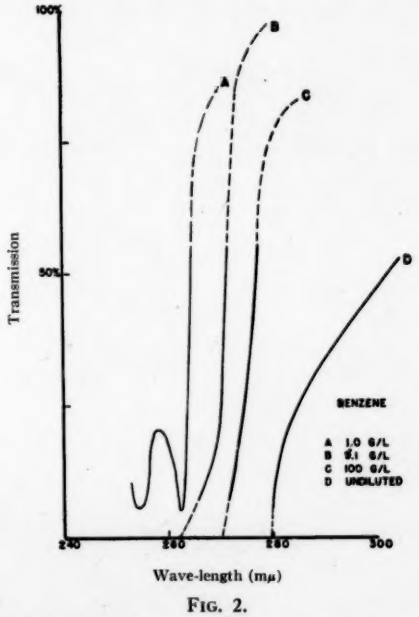


FIG. 2.

* Developed in connection with research supported by the Office of Naval Research.

[1] T. M. Dahm, J. Opt. Soc. Am. and Rev. Sci. Inst. **15**, 266 (1927).
[2] K. S. Gibson, I. C. T. **5**, 271 (1929).
[3] A. L. Hughes and L. A. Dubridge, *Photoelectric Phenomena* (McGraw-Hill Book Company, Inc., New York 1932), p. 456.
[4] See, for example, American Petroleum Institute Research Project 44 at the National Bureau of Standards. Catalog of Ultraviolet Spectrograms.

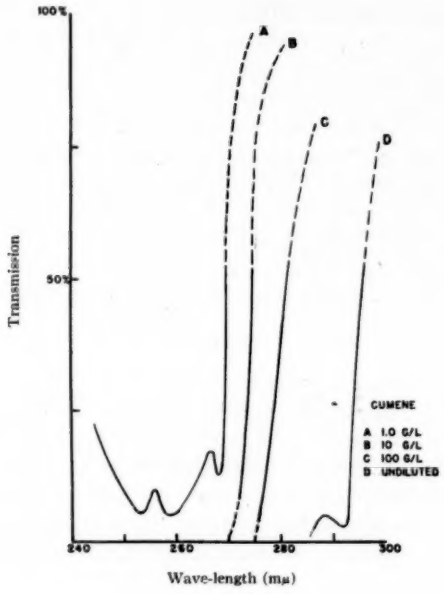


FIG. 3.

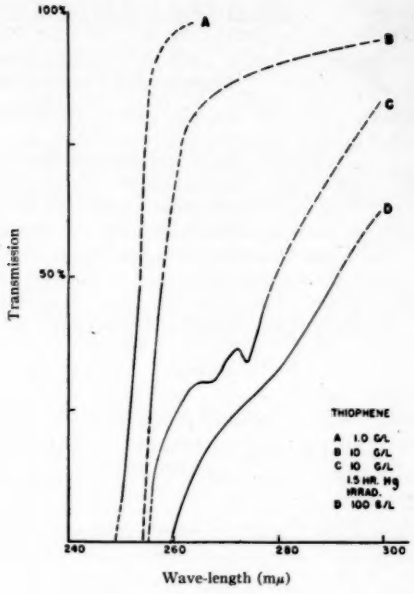


FIG. 5.

especially aromatic substances, absorb more or less strongly in the near ultraviolet and have a fairly sharp cut-off wave-length. These substances were further investigated as to their possible use as filters, and the results are given below.

APPARATUS AND METHOD

The optical density was measured, on a Beckman DU spectrophotometer, for solutions of different concentration of a number of organic compounds. In each case the substance was dissolved in isooctane, and the percent transmission measured against pure isooctane as the standard. In most cases solutions of three different concentrations were used, these being approximately 1.0, 10.0, and 100 g/l. The isooctane was the best Rohm and Haas grade, purified by drying over NaOH, followed by careful distillation. No special efforts were made to purify the other organic substances, though the best commercially available products were used.

For each of the solutions the percent transmission was measured in the range 2200–3200A. However, it has been found[5] that the response of the Beckman spectrophotometer is linear only in the range of transmission of approximately 1.5 percent to 50 percent, and so the values for transmission lying outside this range were corrected on the basis of Beer's law.

FIG. 4. Absorption of light from Fe arc by thiophene.

(a) Fe arc, no filter.
(b) Fe arc + Corning No. 9863 + thiophene in isooctane 50 g/l.
(c) Fe arc + Corning No. 9863 + thiophene in isooctane 10 g/l.
(d) Fe arc + Corning No. 9863 + thiophene in isooctane 1.0 g/l.
(e) Fe arc + Corning No. 9863 + thiophene in isooctane 0.1 g/l.

[5] J. M. Vandenbelt, J. Forsyth, and A. Garrett, Ind. Eng. Chem., Anal. Ed. **17**, 235 (1945).

RESULTS

The curves reproduced below indicate the results obtained for some of the compounds studied. In these curves the solid line portion indicates measurements made in the region of linearity of the spectrophotometer, while the broken line portions indicate corrected values. Figure 1 shows the transmission of the purified isooctane used as solvent as compared with distilled water. The small amount of structure in the region around 2600A indicates that there were some impurities left in the solvent. But since only relative measurements were made in the other curves, this structure does not affect the shape of those curves. Curves are shown for benzene (see Fig. 2), cumene (see Fig. 3), phenol, and thiophene (see Fig. 5). It is seen that in each case the cut-off is quite sharp, and that the cut-off wave-length is shifted to longer wave-lengths as the concentration is increased. Figure 4 shows a photograph of the iron arc spectrum, and illustrates how the cut-off frequency shifts because of changes in the concentration of a thiophene filter. It should be mentioned that it was intended that these filters be used in conjunction with a glass filter, Corning No. 9863, which transmits only in the region 2350–4000A. In this way the structure that appears at the short wave-length end of some of the spectra is removed in practice.

Some of the compounds studied are unstable under the action of ultraviolet light, and allowance must be made for this in using these substances as filters. Of the compounds here listed thiophene is the most unstable; one of the curves in Fig. 5 shows the change that takes place in the transmission properties of a solution of thiophene in isooctane (10 g/l) after 1.5-hours irradiation by a mercury arc. In using a thiophene filter for long exposures it is thus necessary to change the solution every 20–30 minutes.

By choosing the proper absorbing compound and concentration, these organic filters provide a means for removing light of wave-length shorter than any desired wave-length in the spectral region 2400–3000A.

TABLE I.

Compound	Concentration g/l in isooctane	Cut-off wave-length angstroms
Benzene	1.0	2615*
	9.1	2630
	100	2700
	undiluted	2800
Chlorobenzene	10.0	2750
	100	2810
	undiluted	2875
Cumene	1.0	2670*
	10	2700
	100	2750
	undiluted	2920+
Dihydropyran	0.2	2240+
	1.0	2310
	10	2540+
Mesitylene	1.0	2735*+
	10	2800
	100	2835
	undiluted	2900
Phenol	1.0	2800*
	10	2845+
Thiophene	1.0	2480
	10	2540
	100	2590

In addition to the compounds for which curves are given, a number of other compounds were studied. The cut-off wave-lengths for some of these substances have been collected in Table I. For low concentrations the cut-off wave-length is determined by extrapolating the linear portion of the curves and neglecting the structure that appears at the short wave-length end of the spectrum. Such values are indicated by an asterisk. For high concentrations the cut-off is determined by a similar extrapolation in the cases where the curves do not drop sharply to zero transmission but instead trail off for transmission below 10 percent. Such values are denoted by +.

Notes on the Cos⁴ Law of Illumination

MAX REISS
Hawk-Eye Works, Eastman Kodak Company, Rochester, New York

(Received March 5, 1948)

The formulae for the distribution of illumination in the focal plane of a lens system without and with distortion are derived. These formulae are applied to some extreme wide-angle lenses, and a numerical example is included. Special attention is given to the distortion of the pupils and the importance of stop position for illumination. The formula for the illumination distribution in a monocentric system is derived.

SINCE the publication of the important paper by Slussareff[1] more attention has been given to the problem of illumination in the image plane of an optical system. In a previous paper[2] the present writer has discussed Slussareff's paper, especially in view of his derivation of the size of the entrance pupil based on third-order theory. In a paper presented by Gardner and Washer at the meeting of the Optical Society of America at Cincinnati in 1947 a wide-angle lens with improved illumination was shown, and in a paper by Gardner[3] the illumination produced by a system of finite aperture and the influence of distortion upon the illumination were discussed.

In the present paper the influence of distortion is considered more in detail, and the general formula applied to a special type of wide-angle lens. It has been felt that for the sake of clarity a repetition of some of the formulae might not be superfluous.

I.

The light energy (flux) radiated from a small surface element dA_1 to a small surface element dA_2 (Fig. 1) is given by Lambert's law

$$d\phi = \frac{B \cdot dA_1 \cdot \cos\theta_1 \cdot dA_2 \cdot \cos\theta_2}{r^2}. \tag{1}$$

Here θ_1 and θ_2 are the respective angles between the normal to the surface elements and the line connecting them, and B is the intrinsic brightness of dA_1. The ratio $(dA_1 \cdot \cos\theta_1/r^2)$ or $(dA_2 \cdot \cos\theta_2/r^2)$ can be interpreted as the solid angle subtended by the element dA_1 or dA_2 at the center of dA_2 or dA_1, respectively.

FIG. 1.

Let us assume an optical system which gives a sharp plane image of a plane object. The light flux emanating from a small surface element dA (Fig. 3) within a small solid angle is

$$d\phi = \frac{B \cdot dA \cdot \cos w \cdot dS \cdot \cos w}{r^2}. \tag{2}$$

Here w is the angle between the axis of the pencil and the optical axis, and dS is the area of the cross section of the cone at an arbitrary vertical plane, the distance between object plane and the plane of reference being e. Then from (2) it follows immediately that

$$d\phi = \frac{B \cdot dA \cdot dS \cdot \cos^4 w}{e^2}. \tag{3}$$

If dA' is the image of dA, then since by definition the illumination $dE = d\phi/dA'$, we get the completely general result:

$$dE = B \cdot \frac{dA}{dA'} \cdot \frac{dS \cdot \cos^4 w}{e^2}. \tag{4}$$

If the image is free from distortion, the area-magnification will be constant all over the field and will be equal to the area-magnification for

[1] G. Slussareff, J. Phys. U.S.S.R. **4**, 537 (1941).
[2] M. Reiss, J. Opt. Soc. Am. **35**, 283 (1945).
[3] I. C. Gardner, J. Research Nat. Bur. Stand. **39**, 213 (1947).

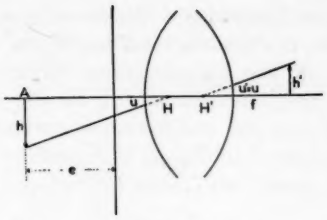


FIG. 2.

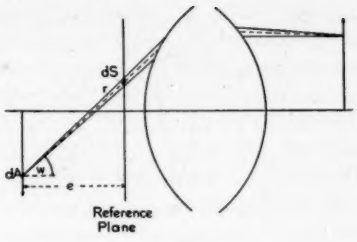


FIG. 3.

the axis, which is the square of the paraxial linear magnification m.

Therefore, assuming freedom from distortion,

$$dE=(B\cdot dS\cdot\cos^4 w/m^2e^2). \tag{5}$$

Up to now we were considering a very small pencil of light. For a finite beam of light we divide the beam into small pencils and get

$$E=\frac{B}{m^2e^2}\int\cos^4 w\cdot dS. \tag{6}$$

The integral has to be taken for the total cross section of the beam at the reference plane. Here w is of course a variable.

II.

If we make the same assumption as in I and let the object travel towards infinity, then $e\cdot m\rightarrow f$, since $m=h'/h$. From the relation of the triangles in Fig. 2 it follows that $h'/h=f/AH$, and for great object distances the difference between AH and e vanishes. We thus have (assuming freedom from distortion)

$$dE=(B\cdot dS\cdot\cos^4 w/f^2). \tag{7}$$

Since for an object point at infinity w does not vary over the aperture, we have for a finite pencil

$$E=(B\cdot S\cdot\cos^4 w/f^2). \tag{8}$$

It may be remembered that S is the area of the cross section of the parallel bundle of rays at a vertical plane in the object space (entrance-pupil).

For a stop in front of the system, S is constant for all obliquities and equal to the area of the stop. For any other stop position we have to determine the actual cross section S of the maximum size bundle which is permitted to enter the lens. The cross section will be a function of the size and position of the stop, the construction of that portion of the lens which lies between the stop and the object space, and of course of any vignetting by the boundaries of the lens. The area generally changes its value with obliquity and it is therefore apparent that, according to (8), the illumination for a system with stop not in front will to some degree depart from the pure $\cos^4$ law. The area S can be determined either photographically or by tracing extreme upper and lower rim rays plus a few of those extreme skew rays which can still go through the system. The intersection of these rays, as they emerge in the object space, with the reference plane gives S.

III.

Assuming the equality of the intrinsic brightness of object and image, the flux emitted from dA and converging at the image is

$$d\phi=\frac{B\cdot dA'\cdot dS'\cdot\cos^4 w'}{e'^2}. \tag{9}$$

Again we have an arbitrary reference plane, calling the cross section of the thin bundle with this plane dS'. (Figure 4.) By dividing (9) by dA' we get the illumination

$$dE=\frac{B\cdot dS'\cdot\cos^4 w'}{e'^2}. \tag{10}$$

For a finite aperture

$$E=\frac{B}{e'^2}\int dS'\cdot\cos^4 w'. \tag{11}$$

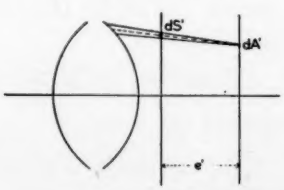


FIG. 4.

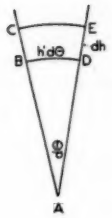


FIG. 5.

By putting (7) equal to (10) we get for the relation between corresponding sizes of entrance and exit pupil

$$\frac{dS'}{dS}=C\frac{\cos^4 w}{\cos^4 w'}. \qquad (12)$$

Since $(\cos^4 w/\cos^4 w')$ is not constant the ratio (12) generally varies with obliquity (distortion of the pupils).

Note that (10) and (11) hold also for a system with distortion. We shall refer to this point later.

IV.

The influence of distortion upon the illumination

It was pointed out that Eq. (4) holds generally even for a system afflicted with distortion, and the problem presents itself of determining the value of dA'/dA for all the conjugate object and image area elements.

For a system with distortion this ratio is not equal to the square of the linear magnification but has to be evaluated as follows (see reference 1):

In Fig. 5 A is the intersection of the optical axis with the image plane, ABC the intersection of the meridional plane with the image plane, and ADE the intersection of another meridional plane with the image plane. The angle between the two meridional planes is $d\theta$. We wish to find the value of the small area $BDEC$. Since $AB=AD=h'=$ the image height, $BC=DE=dh'$, a small increase in height. Then the area is $h'\cdot d\theta\cdot dh'$. The corresponding area in the object plane is $h\cdot d\theta\cdot dh$. From this it follows that

$$dA'/dA=(h'\cdot dh'/h\cdot dh). \qquad (13)$$

By substituting in (4) we get

$$dE=B\cdot\frac{dS\cdot\cos^4 w}{e^2}\cdot\frac{h}{h'}\cdot\frac{dh}{dh'}. \qquad (14)$$

We now introduce a fictitious lens without distortion, of the same focal length and the same position of the Gauss points. Of course the paraxial magnification m will be the same for both the real and the fictitious systems. If the image height of the image of h produced by the fictitious system is h_0', then we replace (14) by

$$dE=\frac{B\cdot dS\cdot\cos^4 w}{e^2}\cdot\frac{hdh}{h_0'dh_0'}\cdot\frac{h_0'dh_0'}{h'dh'}. \qquad (15)$$

Now $hdh/h_0'dh_0'$ is the reciprocal area magnification of the fictitious system which is constant all over the field and equal to the square of the paraxial magnification m_0 of the real system. Therefore

$$dE=\frac{B\cdot dS\cdot\cos^4 w}{e^2m_0^2}\cdot\frac{h_0'dh_0'}{h'dh'}. \qquad (16)$$

If we consider an object at infinity, then again $em_0\rightarrow f$ and

$$dE=\frac{B\cdot dS\cdot\cos^4 w}{f^2}\cdot\frac{h_0'dh_0'}{h'dh'}. \qquad (17)$$

We see that the illumination as compared with a system without distortion is changed by the factor $h_0'dh_0'/h'dh'$.

V.

We can readily see that it is possible to have such an amount of distortion that the illumination will be constant over the entire field (see reference 3). According to (17) the illumination will be constant if $dS\cdot\cos^4 w(h_0'dh_0'/h'dh')$ is constant. If we assume a stop in front of the system then dS is constant for all obliquities and the condition reduces itself to $\cos^4 w(h_0'dh_0'/h'dh')=$ constant over the field. This constant must be unity because the illumination for the axis is known to be $(B\cdot dS/f^2)$. We therefore start from

$$\cos^4 w(h_0'dh_0'/h'dh')=1. \qquad (18)$$

The ideal height is

$$h_0'=f\tan w. \qquad (19)$$

Therefore

$$\cos^4 w \cdot f \cdot \frac{\tan w}{h'(w)} \cdot \frac{f}{\cos^2 w} \cdot \frac{dw}{dh'} = 1, \qquad (20)$$

$$\frac{f^2 \cdot \cos w \cdot \sin w}{h'(w)} \cdot \frac{dw}{dh'} = 1,$$

$$h'(w) \cdot \frac{dh'}{dw} = f^2 \cdot \sin w \cdot \cos w,$$

$$\frac{1}{2} \cdot \frac{dh'^2}{dw} = f^2 \frac{1}{2} \cdot \frac{d \sin^2 w}{dw}.$$

By integration

$$h'^2 = f^2 \sin^2 w + C,$$

$C = 0$ because for $w = 0$, $h' = 0$.

Hence

$$h' = f \sin w. \qquad (21)$$

From (19) and (21) we have

$$h_0' = \frac{h'}{\cos w}.$$

Hence

$$\frac{dh_0'}{dh'} = \frac{\cos w + h' \sin w \cdot \dfrac{dw}{dh'}}{\cos^2 w},$$

and we have from (21)

$$\frac{dw}{dh'} = \frac{1}{f \cos w}.$$

Therefore

$$\frac{dh_0'}{dh'} = \frac{f \cos^2 w + h' \sin w}{f \cos^3 w}.$$

According to (21) $h' \sin w = f \sin^2 w$. Therefore $dh_0'/dh' = 1/\cos^3 w$. This is the ratio of the radial magnification of corresponding points in the undistorted and distorted images for the special case of distortion represented by Eq. (21).

Starting from (16) we get the analogous condition for constant illumination if the object is at a finite distance:

$$h' = em_0 \cdot \sin w. \qquad (22)$$

It can be seen that a system with a stop between lens and focal plane cannot have constant illumination. The illumination in this case falls off proportional to $\cos^4 w'$ for small apertures (10) and according to (11) for any aperture. However, even with a stop in this position the illumination can be improved, as compared with a system free from distortion, by introducing negative (barrel) distortion. This simply means that in the system with distortion a fixed point in the image plane corresponds to a greater field angle in the object space than in the distortion-free system. Actually the lens system discussed in VII has the stop so close to the rear element that the exit pupil is practically constant in position and size.

But if a system is required to have constant illumination over the field then it must have the stop inside or in front of the system. In this case $E = (B \cdot dS' \cdot \cos^4 w'/e'^2)$ will be constant due to the fact that the product $dS' \cdot \cos^4 w'$ is constant. This means that the exit pupil dS' increases its size with increasing w' (distortion of the exit pupil).

VI.

In the lecture by Gardner and Washer[3a] mentioned above, a sample of the "Pleon" lens recently brought from Germany was discussed. The image height of this lens follows the law:

$$h' = f \cdot w, \qquad (23)$$

corresponding to an image height between that of a distortion-free system and the system of V (Fig. 6). This is the type of correction in which objects subtending the same angle at the lens give images of the same size (equidistant projection) (W. N. Bond,[4] Hill[5]). It was already pointed out by Hill[5] that such a lens gives improved illumination as compared with a distortionless lens. The illumination can easily be

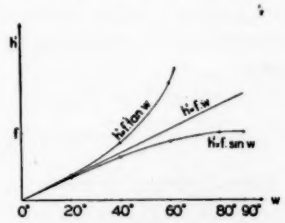


FIG. 6.

[3a] I. C. Gardner and F. E. Washer, J. Opt. Soc. Am. **38**, 421 (1948), published while this paper was in print.

[4] W. N. Bond, Phil. Mag. **44**, 999 (1922).

[5] R. Hill, "A lens for whole sky photography," Proc. Opt. Conv. **2**, 878 (1926).

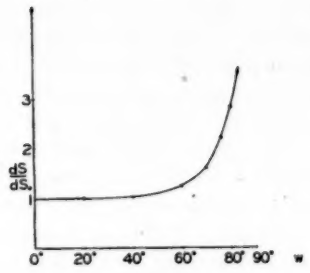


FIG. 6a.

derived as follows: since

$$h' = f \cdot w$$

and

$$h_0' = f \tan w,$$

then dh_0'/dh' can be written $(dh_0'/dw) \cdot (dw/dh')$. Now

$$\frac{dh_0'}{dw} = \frac{f}{\cos^2 w},$$

$$\frac{dw}{dh'} = \frac{1}{f},$$

or

$$\frac{dh_0'}{dh'} = \frac{1}{\cos^2 w},$$

$$\frac{h_0'}{h'} \cdot \frac{dh_0'}{dh'} = \frac{1}{\cos^2 w} \cdot \frac{\tan w}{w} = \frac{\sin w}{w \cdot \cos^3 w}.$$

Substituting in (17) we have

$$dE = \frac{B \cdot dS}{f^2} \cdot \frac{\sin w \cdot \cos w}{w} \qquad (24)$$

(dS generally varies with obliquity). This gives for $w = 0$ an indeterminate value. By differentiating numerator and denominator and putting $w = 0$, we get

$$dE = B \cdot dS / f^2,$$

For $w = 90°$, $h' = \pi/2$, $(\sin w \cos w / w) = 0$. In a perfect system the distortion does not change with a shift of the stop. If the stop of the Pleon were in front of the lens (disregarding the fact that this is impractical) then using (10) we see that in order to satisfy $E = 0$ for $w = 90°$ the exit pupil would shrink to a value of zero. If, on the other hand, the stop is between the lens and the focal plane, then dS' is constant for all obliquities, and $dS' \cdot \cos^4 w'$ cannot become 0. In order to reconcile this with Eq. (24) it must follow that for w approaching 90° the limit of $dS \cdot \cos w$ remains finite. This means that in this case the entrance pupil becomes distorted up to a value of infinity.

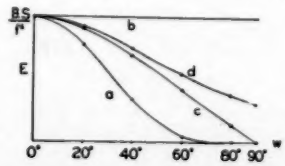


FIG. 7.

In putting (24) equal to (10) we have

$$\frac{B \cdot dS}{f^2} \cdot \frac{\sin w \cos w}{w} = \frac{B \cdot dS' \cdot \cos^4 w'}{e'^2}. \qquad (24a)$$

For the axis:

$$(B \cdot dS_0/f^2) = (B \cdot dS_0'/e'^2). \qquad (24b)$$

If we assume the stop to be between lens and focal plane then dS' is constant for all obliquities, and by dividing (24a) by (24b) we get

$$\frac{dS}{dS_0} = \frac{w}{\sin w \cdot \cos w} \cos^4 w' \qquad (24c)$$

(for $h' = fw$ and constant exit pupil).

According to (24c) the ratio dS/dS_0 also depends upon the angle between the principal ray in the image space and the axis. This angle depends on the stop position. The lens described by R. Hill in British patent 225,398 is in its construction similar to the Pleon and very closely follows the law $h' = fw$, and has the stop very close to the rear element so that the position and size of the exit pupil are nearly constant. Figure 6a gives the ratio dS/dS_0, showing the rapid increase of the size of the entrance pupil from 60° upwards. Figure 7 shows the illumination characteristics for:

(a) a distortionless lens with stop in front (strict $\cos^4 w$ law),
(b) a system where $h' = f \cdot \sin w$ with stop in front (illumination constant over the entire field),
(c) a system where $h' = f \cdot w$ with stop in front (illumination following Eq. (24)),

(d) a system where $h'=f\cdot w$ with stop between lens and focal plane, (illumination following (24) with dS increasing with obliquity. Illumination strictly follows (10) for small apertures. We notice the improvement of the illumination by shifting the stop (c to d).

The general case of a system with distortion

The general equation for any system with distortion is (17),

$$dE=\frac{B\cdot dS\cdot\cos^4 w}{f^2}\cdot\frac{h_0'dh_0'}{h'dh'}.$$

Here

$$h_0'=f\cdot\tan w;\quad h'=\Phi(w);\quad \frac{dh_0'}{dh'}=\frac{dh_0'}{dw}\cdot\frac{dw}{dh'},$$

therefore

$$\frac{h_0'dh_0'}{h'dh'}=\frac{f\tan w}{h'(w)}\cdot\frac{f}{\cos^2 w\dfrac{dh'}{dw}}=\frac{f^2\sin w}{\cos^3 w\cdot h'(w)\dfrac{dh'}{dw}}.$$

After substituting in (17) we get

$$dE=\frac{B\cdot dS\cdot\sin w\cdot\cos w}{h'(w)\dfrac{dh'}{dw}}. \qquad (25)$$

If $h'(w)$ is given in analytical form as in the last two examples, the solution can immediately be given. But for an arbitrary system with distortion this will not be the case. As an example, the illumination was computed for a particular lens which, like the Pleon, is a modification of the Hill Sky-Lens.[5] The specifications of the lens are as follows (U. S. 2,126,126 Merte): ($f=100$ mm)

r_1 = plano,	d_1 =	6.55,	
r_2 = 215.58,	b_1 =	110.28,	
r_3 = −42.70,	b_2 =	2.36,	
r_4 = −21.20,	d_2 =	6.81,	
r_5 = −16.15,	l =	1.31,	
r_6 = 100.88,	d_3 =	2.99,	
r_7 = −23.08,	d_4 =	6.42,	
L_1 = 1.62313	57.0,		
L_2 = 1.51746	58.9,		
L_3 = 1.55088	49.6,		
L_4 = 1.60284	60.6.		

It was found that this lens followed neither of the laws $h'=f\sin w$ nor $h'=fw$.

For the determination of the illumination at $h'(w)$ we chose an image point below and above h' and connected them by a curve

$$h'=aw+bw^2+cw^3.$$

The a, b, and c were computed by solving the three linear equations after putting in the three corresponding pairs of h' and w obtained by ray tracing.

After having the analytical connection of h' and w in the neighborhood of the particular h' in which we were interested, we obtained

$$\frac{dh'}{dw}=a+2bw+3cw^2.$$

This was done for an angle $w=61°7113$, giving $h'=9.511729$ and $h_0'=18.458027$, with the result:

$$h'=-9.362974+1.666816w^2+2.005904w^3.$$

The size of the entrance pupil was then determined by tracing extreme upper and lower rim rays and the two extreme sagittal rays corresponding to an f:16 beam. Through the given four points (intersections of the rays with the reference plane in the object space) we then laid an ellipse, which was found to have an area of 0.262951 sq. cm.[6] This finally gives, for $w=61°7113$:

$$E=0.001930B.$$

For the axis

$$E=0.003118B.$$

This means that for the present system the illumination at 61° is still 62 percent of that on the axis. For a system without distortion it would only be 5 percent.

The area of the exit pupil projected on to a plane in contact with the vertex of the last surface was then computed by using the intersection points of the four chosen rays with that plane. This gave $dS'=0.785398$, which is almost exactly the value of the exit pupil for the axial beam.

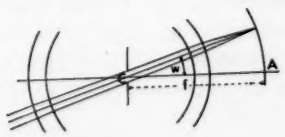


FIG. 8.

[6] The entrance pupil of the axial beam is 0.302752 (distortion of the entrance pupil of the oblique beam).

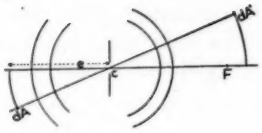


FIG. 9.

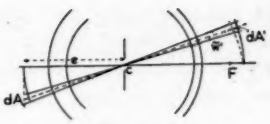


FIG. 10.

Using the formula[7]

$$dE=\frac{B\cdot dS'\cdot\cos^4 w}{e'^2}, \tag{10}$$

this gives

$$E=0.001932B,$$
$$w=61°.$$

This is a striking illustration of the soundness of the formulae.

VIII.

The monocentric case

In a monocentric system all surfaces have a common center. Therefore, every ray going through the center is unrefracted and every principal ray is an optical axis[8] (Fig. 8).

For an object at infinity the image must lie on a sphere with its center at C. By reference to the image space situation, we can see without any further computation that the illumination will be proportional to $\cos w$. For, since the image element is perpendicular to the principal ray, one $\cos w'$ is omitted. The stop is situated at the Gauss points and for small stop openings the exit pupil will be coincident with stop opening in position and size. The distance between the image element and C is f and $w'=w$. We therefore have, according to (1),

$$dE=(B\cdot dS\cdot\cos w/f^2). \tag{26}$$

If we start our consideration in the object space the result is not so obvious (Fig. 9). Let us first consider a spherical object (center at C). For reasons of symmetry this will give a spherical image. Since dA is perpendicular to the principal ray the illumination is (see (1))

$$dE=\frac{B\cdot dA}{dA'}\frac{dS\cdot\cos w}{e^2},$$

$$dE=B(dS\cdot\cos w/m^2e^2).$$

If the object travels towards infinity we have

$$dE=(B\cdot dS\cdot\cos w/f^2), \tag{27}$$

where S is the entrance pupil coinciding with stop.

Although it might seem superfluous we can derive the same result by assuming a flat object and let it travel towards infinity. Since a sphere is imaged as a sphere a plane with the same vertex will be imaged on a surface departing from a sphere (see Fig. 10). The illumination on the image of the plane will be

$$dE=\frac{B\cdot dS\cdot\cos^4 w}{(dA'/dA)e^2}. \tag{28}$$

The solid angle subtended by dA at C is $(dA\cdot\cos^3 w/e^2)$. The same solid angle is subtended by dA' at $C(dA'\cdot\cos w'^*/b^2)$, where b is the distance of dA' from C, and w'^* is the angle between the principal ray and the normal at dA'. Therefore

$$(dA'/dA)=(\cos^3 w/e^2)\cdot(b^2/\cos w'^*).$$

As the plane object travels towards infinity $w'^*\to 0$ and $b\to f$, or

$$(dA'/dA)=(\cos^3 w\cdot f^2/e^2).$$

Substituting this in (28), we get

$$dE=(B\cdot dS\cdot\cos w/f^2). \tag{29}$$

For a finite bundle

$$E=B\cdot S\cdot(\cos w/f^2). \tag{30}$$

Because of the symmetrical type of this lens the entrance pupil will be constant for all obliquities except for departures of smaller order.

In conclusion I should like to thank Professor R. Kingslake for his interest in this paper and for many valuable discussions.

[7] We should strictly use the formula $E=(B/e'^2)\int dS'\cdot\cos^4 w'$, but as Gardner has shown (see reference 3) this makes little difference for small apertures.

[8] Obviously the Gauss points coincide with the center C and the distance $AC=f$.

Letters to the Editor

Elimination of Water Vapor in Infra-Red Spectrometers*

PAUL A. GIGUÈRE AND RICHARD M. BADGER
Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California
August 13, 1948

ATMOSPHERIC water vapor and carbon dioxide cause serious difficulties in infra-red spectroscopy. In particular, the fundamental bands of water at 2.7 and 6.3μ are relatively very intense; the latter of these frequently makes observation impossible or quite unreliable in a rather broad region of the spectrum which is sometimes of considerable interest. The most effective way of eliminating the contamination is to evacuate the spectrometer proper and as much of the external path as possible. This has been done in a number of recently constructed instruments.[1,2] A large automatic recording spectrometer built in these laboratories[3] can readily be evacuated to a pressure of about 0.1 mm with the use only of a mechanical pump, which eliminates the atmospheric bands almost entirely. Unfortunately, the vacuum construction usually requires a rather bulky and costly instrument. Another partial solution to the problem is the double beam construction which has been incorporated into some recent instruments. For this arrangement to be completely effective, however, the two optical paths must be equivalent, which is often not realized in practice; also the resolving power of the spectrometer must be great in comparison with the width of the water-vapor lines. Since the majority of commercially available spectrometers are not evacuable and do not incorporate the double beam feature, other means must be used to remove the contaminants. Little quantitative information exists in the literature which can be helpful in this connection, and it is the purpose of the present note to report experience of our own which may be of some assistance to other investigators.

As has been observed before, we have found the use of drying agents within the spectrometer case surprisingly ineffective. This result has sometimes been attributed to occlusion of water in the rocksalt prisms and windows or to adsorption on them. If such were the case, the absorption bands should show the characteristics of a liquid and not of a gas, as is observed. A more plausible explanation lies in the fact that desiccation depending entirely on diffusion processes is extremely slow and easily nullified by "breathing" of the instrument as a result of changes in atmospheric pressure and temperature. A more effective use of drying agents has been made by Smith[4] who succeeded in drying a large grating spectrograph by continuously circulating through it a stream of air which had passed over trays of P_2O_5 stacked in separate boxes. The drying agent could be renewed periodically without opening the spectrograph.

Our own experience relates particularly to a Beckman infra-red spectrometer, Model IR-2. Since the use of desiccants was unsatisfactory, we resorted to the flushing of all parts of the optical path with a dry gas. It has been recommended to use for this purpose the gas escaping from a container of liquid nitrogen, but we have found that commercial "dry" nitrogen from a cylinder is more convenient and equally satisfactory. First the monochromator is flushed for a few minutes with a moderate stream which is then reduced to a steady flow of about ten liters per hour, or enough to maintain a barely measurable pressure at the openings in the spectrometer case. Similar attention must be given to all parts of the optical path outside the monochromator. In particular, the compartment holding the thermocouple has poor communication with the adjacent chamber containing drying agent, and it should either be provided with separate desiccant or flushed with nitrogen. The auxiliary absorption cell (27-cm path) may either be evacuated or the window opposite to the Nernst glower removed and a stream of dry gas introduced through the inlet tube. Care must be taken that the flow of gas be gentle and steady so as not to affect the glower. When the 27-cm cell is not used, there appears to be some advantage in removing the window which otherwise may collect a deposit sublimed from the Nernst glower. A very serious "ghost" absorption at 7.3μ developed in our instrument after a few weeks of operation which was finally traced to this cause. The short compartment designed for the liquid absorption cell should likewise be flushed with nitrogen for a few minutes before the instrument is to be used. The full effect of each drying operation is attained surprisingly rapidly (10–15 min.).

We have not been able to find in the literature any quantitative data on the intensity of the infra-red bands of water. Therefore, we believe that the results presented below may be of some use to others, for instance, in estimating the efficiency of a given drying procedure. Curve *A* of Fig. 1 shows a tracing of part of the 6.3-μ band with the instrument dried as described above. By admitting laboratory air (63 percent humidity at 25°C) into various sections of the spectrometer curves *B*, *C*, and *D* were obtained. Beer's law was applied to these data as follows: the intensity of transmitted radiation I was measured at 6.4μ, which is about the maximum of the ν_2 band of water.

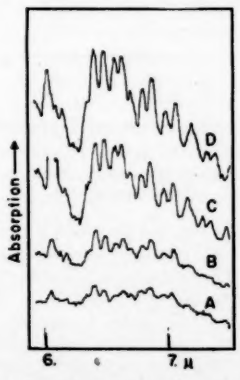


FIG. 1. The water vapor band at 6.3μ. *A*: Residual absorption in the spectrometer dried as described. *B*: 5.5-cm air path. *C*: 27-cm air path. *D*: 32.5-cm air path at 63 percent humidity and 25°C.

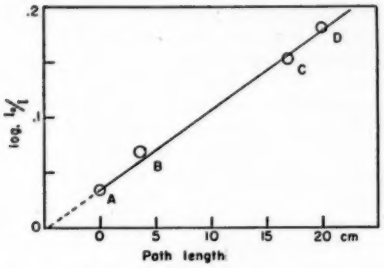


FIG. 2. Extinction of radiation at 6.4μ by water vapor, as a function of path length in saturated air at 25°C. The letters correspond to the curves in Fig. 1.

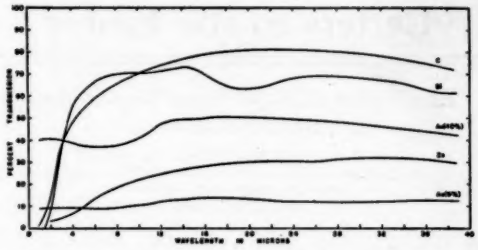


FIG. 1. The transmission of films of carbon, bismuth, zinc, and gold blacks. The transmission values are corrected for the radiation losses of the supporting window. A KBr prism was used from 15 to 24 μ and a thallium bromide-iodide prism from 24 to 39 μ.

Since it was not possible to remove entirely all traces of water vapor in the instrument, the value of I_0 could not be determined at exactly the same wave-length; instead, it was measured at 6.25μ, which corresponds to a minimum of absorption in that band. The values thus obtained for I_0 did not change by more than 2–3 percent on going from curve *A* to *D*, an indication that there is very little absorption at that particular frequency. This was further confirmed by the fact that the data fall quite closely on a straight line, Fig. 2. Extrapolation of this curve to zero absorption indicates that, even with all the above precautions, the amount of water vapor remaining in the optical path was still equivalent to nearly five cm of saturated air. As the total path length was of the order of two meters, the residual humidity was therefore 2.5 percent. This last amount of water vapor could probably not be removed easily. Experience has shown that no noticeable improvement results from passing the "dry" nitrogen over trains of P_2O_5 or other desiccants. The use of three gas traps in series, filled with silica gel and immersed in liquid air, reduced absorption only slightly.

* Contribution No. 1235.
[1] E. E. Bell, R. H. Noble, and H. H. Nielsen, Rev. Sci. Inst. **18**, 48 (1947).
[2] E. J. Martin, A. W. Fischer, B. Mandel, and R. E. Nusbaum, J. Opt. Soc. Am. **37**, 923 (1947).
[3] R. M. Badger, L. R. Zumwalt, and P. A. Giguère, Rev. Sci. Inst., to be published.
[4] L. G. Smith, Rev. Sci. Inst. **13**, 54 (1942).

Infra-Red Absorption of Deposited Blacks

EARLE K. PLYLER AND JOSEPH J. BALL
National Bureau of Standards, Washington, D. C.
September 20, 1948

RECENTLY there has appeared a report on the preparation and optical properties of gold blacks.[1] It was found that the conditions of depositing the films greatly influenced the infra-red absorption of the films.

Professor Harris has kindly furnished the writer with three films of gold black deposited on windows of thallium bromide-iodide (KRS-5). This has made it possible to extend the absorption measurements to 39 μ. It has been shown by Professor Harris that films prepared in an atmosphere of nitrogen with only a trace of oxygen present are best suited for infra-red absorption in thin layers. A discussion of the methods of deposition and the infra-red absorption to 15 μ are given in his paper.[1]

In Fig. 1 is shown the infra-red transmission of carbon, bismuth, zinc, and gold blacks. The carbon black was deposited on the thallium bromide-iodide window by placing it above burning camphor gum. By constantly moving the window above the flame, a smooth coating was obtained which transmitted about 0.1 percent in the visible spectrum. The bismuth black and zinc black were deposited by Mr. Weyrick at the University of Michigan, using the Pfund method.[2] The bismuth black transmitted about one percent of the visible radiant energy, and the zinc black transmitted two percent.

The uncoated thallium bromide-iodide window transmitted about 70 percent of the incident energy in the 2 to 15-μ region, and the results in Fig. 1 have been corrected for this loss. The three samples of gold blacks showed 1, 9, and 40 percent transmission in the visible region. The heavier coating did not transmit over four percent at any wave-length from 1 to 39 μ and the graph is omitted.

It can be seen that the gold blacks do not change appreciably in transmission in comparing the longer wave-length region to the shorter wave-length region which is in contrast to the carbon, bismuth, and zinc blacks. Harris, McGinnies, and Siegel[1] found that, under certain conditions of depositing the gold black, transmission increased rapidly beyond 1 μ. It may be possible that the bismuth and zinc blacks would not become so transparent in the long wave-length region if the deposit was made under different conditions.

The reflection of the gold black was measured from 1 to 39 μ, and it did not amount to over one percent in any part of the region. The film was smooth and regular reflection could be observed in the visible region for incident angles greater than 70°. For thermocouples with a rapid response a thin film of material which is highly absorbing is necessary. These measurements show that gold black is well suited for coating receivers. Further measurements will be necessary to determine the absorption of these films beyond 39 microns.

[1] L. Harris, R. T. McGinnies, and B. M. Siegel, J. Opt. Soc. Am. **38**, 582 (1948).
[2] A. H. Pfund, J. Opt. Soc. Am. **23**, 375 (1933).

The Emissivity of Globar*

SHIRLEIGH SILVERMAN
Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland
September 17, 1948

RECENTLY the writer has been interested in flame temperature measurements, utilizing the sort of infra-red brightness thermometer first described by Schmidt[1] and later by Henning and Tingwaldt.[2] This type of thermometry becomes relatively simple, accurate, and quite rapid, provided a good comparison radiation standard is available. Since a standard blackbody is somewhat of a nuisance to prepare and maintain, it is much more convenient to utilize a common heating element such as a globar for the radiation standard. However, the literature is devoid of good data on the absolute emissivity of globar burners. Accordingly, the writer felt it worth while to obtain the emissivity of globar in the region 1.5–15 microns.

The emissivity of globar relative to that of a blackbody was obtained in the following manner. A tubular globar element 15″ long, 1¼″ O.D., wall thickness $\frac{5}{16}$″ was purchased from the Carborundum Company. A slot about 2″ long by 1.0 mm wide was cut into the center portion of one side of the element. The radiation from the slot closely approximates that of a blackbody, and may be readily compared with that from a surface area immediately adjacent to the slot. The tubular element is preferred to a solid rod, since it minimizes the temperature gradient between inner and outer portions of the globar. The element was placed in a water-cooled jacket with a 3″ × 1.5″ open window for radiation measurements and run at 1375° K. Readings were taken at various wave-lengths on a Perkin-Elmer 12-C spectrometer by focusing the "blackbody" slot and globar surface alternately on the slit of the spectrometer. The emissivity is shown in Fig. 1. It is rather interesting to note that the small maximum at around 8 to 10 microns and the minimum at about 12.5 microns lie, respectively, as one would expect, at the wave-lengths of minimum and maximum reflectivity of crystalline carborundum[3] as shown in Fig. 1.

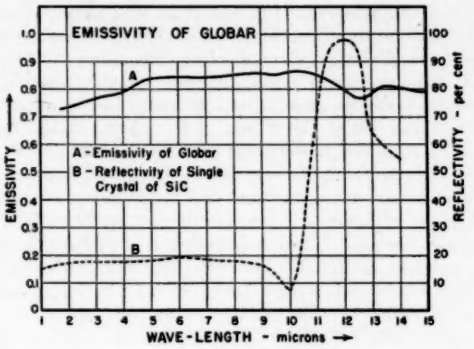


FIG. 1.

In conclusion, the author is much indebted to Mr. B. A. Bovee of the Carborundum Company for his kindness in preparing the special globar used in this experiment.

* This work was supported by the Bureau of Ordnance, Navy Department under Contract NOrd-7386.
[1] H. Schmidt, Ann. d. Physik **29**, 1027 (1909).
[2] F. Henning and C. Tingwaldt, Zeits. f. Physik **48**, 805 (1928).
[3] Coblentz, *Supplementary Investigations of Infrared Spectra V* (Carnegie Institution of Washington, 1909), p. 36.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 11 NOVEMBER, 1948

Book Review

Symposium on Searchlights

Pp. 164, Figs. 116. The ILLUMINATING ENGINEERING SOCIETY, London, 1948.

The title page carries the statement "containing a series of five papers presented at a meeting of the Illuminating Engineering Society, held in London on April 15, 1947, the discussion on those papers and a further written contribution." Even if these six papers cover most of the investigative work done in England during World War II in connection with the development of searchlights for military applications, a cursory review reveals that probably as much and perhaps more was accomplished than in any other country, certainly more than in the United States.

In the paper on "The function and design of army searchlights," by E. W. Chivers and D. E. H. Jones, one finds evidence to counter a possible assumption that the carbon-arc searchlight has given way completely to radar for anti-aircraft defense. It is pointed out that for radar direction of searchlights during bad atmospheric conditions ". . . the provision of an intersection of three beams at the target may provide an adequate indication for a fighter pilot to make contact with the attacking aircraft." Of particular interest is the discussion and accompanying photographs on reflector flash tests. The reviewer has also found this test useful for determining reflector imperfections and for optimum focusing of the source for beam candlepower measurements. For those interested in beam intensity *versus* anti-aircraft visual range, the test results presented are gratifying. They agree in the main with several of the theoretical predictions which have been made.

More comprehensive discussions on "seeing" by means of a searchlight are given in the two papers: "The visibility of targets in a naval searchlight beam," by W. D. Chesterman and W. S. Stiles and "Some visibility problems associated with anti-aircraft searchlight beams," by S. S. Beggs

and J. M. Waldram. The first deals with the brightness distribution of the searchlight beam as contrasted with that of the marine target illuminated, the laboratory simulation of these computed conditions and visual observations thereof, and the subsequent calculation of the range at which the actual target should be clearly visible. Empirical equations and several range tables are presented showing the results of varying the beam width and peak candlepower, the target size and reflectivity, the offset distance of observer from searchlight, and the atmospheric attenuation. The second of these two papers tackles the visibility of aircraft. As explained, the aircraft case is more difficult; the ranges are greater, scattering of light by the atmosphere varies with altitude, and the geometry is three- rather than two-dimensional. Three problems are considered; targets seen from the ground, targets seen from the air at about the same altitude, and the visibility of the searchlight beam from the air. The results for three measured atmospheric conditions are given in graphical form. In connection with the first problem, the discussion on the perspective geometry of the "end of the beam" is most revealing. This is the first time the reviewer has seen a discussion of this phenomenon.

The paper on "Photometry of searchlights," by H. K. Cameron, E. H. Rayner, E. R. Thomas, and G. T. Winch should be of especial interest to all who are concerned with photometric measurements of projectors. Because of the wartime blackout conditions, photoelectric telephotometric methods were devised for making beam intensity measurements outdoors during daylight hours. This information will be most welcome to those who have been forced to make such measurements at night. Of fundamental importance is the finding that "measurements of beam intensity distribution made over photometric ranges of 1500 to 5000 ft. confirm theoretical calculations that, provided the projector is focused for optimum flash at the range at which measurements are made and not to infinity, no significant differences in beam intensity or distribution are observed at the different distances." By "optimum flash," the above-mentioned mirror flash test is used and an ingenious extension of the technique is described whereby concentric rings of differently dyed gelatin filters are mounted upon an illuminated opal glass surface for use as a disk source at the focus of the parabolic mirror. Visual viewing or color photographs make possible a correlation of flashed zones of the reflector with those of the source. For those concerned with the brightness distribution of non-steady sources, the rapid-scan brightness measuring apparatus described will be stimulating.

The other two papers, "High powered searchlights capable of wide divergence," by W. Helmore, H. K. Cameron, F. S. Hawkins, L. B. W. Jolley, and L. M. King-Brewster and "Aircraft searchlights for anti-submarine warfare," by C. J. Carr, are concerned with the development of high candlepower projectors for employment on aircraft. Startling operating parameters for carbon-arc lamps are presented in the first of these two papers. Power inputs up to 620 kw, arc currents of 4000 amperes, positive cored carbons of 64-mm diameter, and disk negatives are some of the surprising details mentioned in connection with a project which initially applied to the development of a light for illuminating enemy bombers from pursuing fighter aircraft. The last paper discusses the development of airborne searchlights used against submarines. This is the one searchlight development reported at the symposium that was employed extensively and effectively during the war. Radar in the airplane detected the surfaced U-boats, and the searchlight provided the illumination for the night attacks. The 1000-pound retractable and 650-pound bomb-rack, wing-mounted, carbon-arc Leigh Light installations and experimental 200-pound "Pumpkin" lights using 1000-watt, water-cooled Hg arcs or 2.5-kw tungsten filament lamps are described.

Discussions of all the papers except the one on antisubmarine searchlights adds to the value of the book in bringing out additional information not contained in the original presentations. This book should be on the "must be read" list of everyone engaged in military searchlight development and applications.

MAURICE K. LAUFER,
Engineer Research and Development Laboratories, Fort Belvoir, Virginia